STIC-Biotech/ChemLib 100781

Fr m: Baum, Stuart
Sent: Friday, August 08, 2003 4:35 PM
To: STIC-Biotech/ChemLib
Subject: Sequence Search....thank you

Please do a standard and interference sequence search for:

(1) SEQ ID NO:78
(2) DNA's that encode SEQ ID NO:78

EXAMINER # 78896
DATE:8/8/2003

SERIAL NUMBER:09/921992

MAIL BOX ROOM: CM1 9E12
OFFICE ROOM # CM1 9D06

RESULTS FORMAT: PAPER

THANKS...Stuart

Stuart F. Baum Ph.D
Art Unit 1638
CM1 9D06
(703) 305-6997
stuart.baum@uspto.gov

Edward Hart
Technical Info. Specialist
STIC/Biotech
CMI 6B02 Tel: 305-9203

Searcher:_____
Phone:_____
Location:_____
Date Picked Up: 8/8/03
Date Completed: 8/8/03
Searcher Prep/Review:_____
Clerical:_____
Online time:_____

TYPE OF SEARCH:
NA Sequences: 1+1
AA Sequences:_____
Structures:_____
Bibliographic:_____
Litigation:_____
Full text:_____
Patent Family:_____
Other:_____

VENDOR/COST (where applic.)
STN:_____
DIALOG:_____
Questel/Orbit:_____
DRLink:_____
Lexis/Nexis:_____
Sequence Sys.: Oligos
WWW/Internet:_____
Other (specify):_____

STIC Search Report
Biotech-Chem Library

STIC Database Tracking Number: 100781

TO: Stuart Baum **Location: CM-1/9D06/9E12** **Art Unit: 1638** **Friday, August 15, 2003** **Case Serial Number: 09921992**	**From: Edward Hart** **Location: Biotech-Chem Library** **CM1-6B02** **Phone: 305-9203** **edward.hart@uspto.gov**

Search Notes

Examiner Baum,

Here are the results of the search you requested.

Please feel free to contact me if you have any questions.

Edward Hart

STIC SEARCH RESULTS

Biotech-Chem Library

Questions about the scope or the results of the search? Contact ***the searcher or contact:***

Mary Hale, Information Branch Supervisor
308-4258, CM1-1E01

Voluntary Results Feedback Form

- *I am an examiner in Workgroup:* ☐ *Example: 1610*
- *Relevant prior art **found**, search results used as follows:*
  - ☐ 102 rejection
  - ☐ 103 rejection
  - ☐ Cited as being of interest.
  - ☐ Helped examiner better understand the invention.
  - ☐ Helped examiner better understand the state of the art in their technology.

Types of relevant prior art found:

- ☐ Foreign Patent(s)
- ☐ Non-Patent Literature
  (journal articles, conference proceedings, new product announcements etc.)

- *Relevant prior art **not found:***
  - ☐ Results verified the lack of relevant prior art (helped determine patentability).
  - ☐ Results were not useful in determining patentability or understanding the invention.

Comments:

Drop off or send completed forms to STIC/Biotech-Chem Library, CM1 – Circ Desk

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 12, 2003, 09:53:55 ; Search time 85 Seconds
(without alignments)
694.662 Million cell updates/sec

Title: US-09-921-992-78
Perfect score: 1866
Sequence: 1 MHNQAPIQRRKSTRIYVGNV.....RAKASQLDEARRIDVQQVEK 372

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1866	100.0	372	22	AAB45692	E. coli gcpE prote
2	1866	100.0	372	23	AAE19653	Escherichia coli G
3	868	46.5	368	23	ABB47958	Listeria monocytog
4	814.5	43.6	403	23	ABP65322	Bifidobacterium lo
5	794	42.6	359	19	AAW98298	H. pylori GHPO 76
6	785.5	42.1	415	22	AAU56212	Propionibacterium
7	761	40.8	378	22	AAG91951	C glutamicum prote
8	751.5	40.3	387	22	AAG81205	Mycobacterium tube
9	751.5	40.3	387	24	ABP57500	Mycobacterium tube
10	534	28.6	421	24	ABP78786	N. gonorrhoeae ami
11	490	26.3	611	20	AAY37164	Amino acid sequenc
12	470.5	25.2	621	20	AAY34971	Amino acid sequenc
13	462	24.8	740	23	AAE19650	Arabidopsis thalia
14	460	24.7	679	22	AAB45693	P. falciparum gcpE
15	446	23.9	752	21	AAW90883	P. falciparum gcpE
16	445.5	23.9	603	23	AAE19652	Oryza sativa GCPE
17	445.5	23.9	686	23	AAE19651	Oryza sativa GCPE
18	428.5	23.0	169	21	AAB41371	Human ORFX ORF1135
19	428.5	23.0	169	23	ABP32470	Human ORF1443 prot
20	398	21.3	144	23	ABU52039	Helicobacter pylor
21	242	13.0	776	22	ABG25879	Novel human diagno
22	119	6.4	113	21	AAG08755	Arabidopsis thalia
23	119	6.4	113	21	AAG45732	Arabidopsis thalia
24	113	6.1	489	22	AAB96357	Putative P. abyssi
25	111.5	6.0	578	20	AAW89273	Granulocytic Ehrli
26	110	5.9	108	21	AAG08756	Arabidopsis thalia
27	110	5.9	108	21	AAG45733	Arabidopsis thalia
28	109	5.8	256	22	AAU53170	Propionibacterium
29	106.5	5.7	284	23	ABB53922	Lactococcus lactis
30	105.5	5.7	578	19	AAW82635	Ehrlichia sp. exte
31	105.5	5.7	578	21	AAY78566	Extended Ehrlichia
32	105.5	5.7	578	23	ABG93408	Ehrlichia antigen
33	105.5	5.7	578	23	AAU73205	Human granulocytic
34	105	5.6	545	20	AAY26187	Soybean 2-isopropy
35	104	5.6	606	23	ABB48094	Listeria monocytog
36	102.5	5.5	542	23	ABB53701	Lactococcus lactis
37	102	5.5	546	23	ABP26321	Streptococcus poly
38	102	5.5	546	23	ABP28474	Streptococcus poly
39	101	5.4	338	23	ABB49213	Listeria monocytog
40	101	5.4	523	22	AAB61619	Alpha-isopropylmal
41	100	5.4	343	19	AAW71467	Cercospora nicotia
42	100	5.4	343	21	AAY99880	Cercospora nicotia
43	100	5.4	502	22	AAG99943	ERA binding domain
44	99	5.3	541	21	AAY91970	PMWaV-2 heat shock
45	99	5.3	2368	22	AAU34139	Staphylococcus aur

ALIGNMENTS

RESULT 1
AAB45692
ID AAB45692 standard; Protein; 372 AA.
XX
AC AAB45692;
XX
DT 15-MAR-2001 (first entry)
XX
DE E. coli gcpE protein.
XX
KW Isoprenoid; gcpE; yfgB; antimicrobial; transgenic plant; agriculture;
KW antimycotic; antiparasitic; antiviral; fungicidal; herbicidal.
XX
OS Escherichia coli.
XX
PN WO200072022-A1.
XX
PD 30-NOV-2000.
XX
PF 20-MAY-2000; 2000WO-EP04592.
XX
PR 21-MAY-1999; 99DE-1023567.
PR 21-MAY-1999; 99DE-1023568.
XX
PA (JOMA/) JOMAA H.
XX
PI Jomaa H;
XX
DR WPI; 2001-025196/03.
DR N-PSDB; AAC82653.
XX
PT Incorporating gcpE and yfgB genes into viruses and cells, for

PT increasing isoprenoid content and identifying e.g. antimicrobial
PT agents, comprises using DNA sequences from bacteria or parasites -
XX
PS Disclosure; Page 15-17; 36pp; German.
XX
CC This invention describes a novel method for incorporating gcpE and yfgB
CC genes into viruses and cells for increasing isoprenoid content and
CC identifying e.g. antimicrobial agents, comprises using DNA sequences (I)
CC from the gcpE or yfgB genes of bacteria or parasites or DNA sequences
CC (II) which hybridize to the specified genes or encode a plastid protein
CC with the same biological activity as those encoded by the genes. The
CC invention also describes (1) plant cells containing (I) or (II); (2)
CC transformed plant cells, and transgenic plants regenerated from them,
CC that contain (I) or (II); (3) determining the enzymatic activity of a
CC gcpE protein; or (4) screening compounds (A) that have antimycotic,
CC antiparasitic or antiviral activity in humans or animals or antiviral,
CC antiparasitic, fungicidal or herbicidal activity in plants. (I) and (II)
CC are used: (i) to increase the isoprenoid levels in viruses and cells;
CC (ii) for determining the enzymatic activity of gcpE and yfgB proteins;
CC and (iii) to identify compounds that inhibit activity of gcpE, i.e.
CC potential antibacterial, antimycotic, antiparasitic or antiviral agents
CC for use in humans or animals, or antiviral, antiparasitic, fungicidal
CC or herbicidal agents for agriculture.
XX
SQ Sequence 372 AA;

Query Match 100.0%; Score 1866; DB 22; Length 372;
Best Local Similarity 100.0%; Pred. No. 7.6e-181;
Matches 372; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MHNQAPIQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MHNQAPIQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVR 60

Qy         61 VSVPTMDAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRM 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VSVPTMDAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRM 120

Qy        121 VVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSV 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 VVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSV 180

Qy        181 KASDVFLAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 KASDVFLAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSL 240

Qy        241 AADPVEEIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AADPVEEIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDV 300

Qy        301 SIIGCVVNGPGEALVSTLGVTGGNKKSGLYEDGVRKDRLDNNDMIDQLEARIRAKASQLD 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SIIGCVVNGPGEALVSTLGVTGGNKKSGLYEDGVRKDRLDNNDMIDQLEARIRAKASQLD 360

Qy        361 EARRIDVQQVEK 372
              ||||||||||||
Db        361 EARRIDVQQVEK 372
```

RESULT 2
AAE19653
ID AAE19653 standard; Protein; 372 AA.
XX
AC AAE19653;
XX
DT 31-MAY-2002 (first entry)
XX
DE Escherichia coli GCPE protein.
XX
KW gcpE gene; methyl-D-erythritol phosphate; MEP; dimethylallyl diphosphate;
KW transgenic plant; isoprenoid compound; tocopherol; isopentyl diphosphate;
KW food; feed source; transfection; single nucleotide polymorphism; SNP;
KW oxidative stress tolerance; UV tolerance; transformation; GCPE protein;
KW plant.
XX
OS Escherichia coli.
XX
PN WO200212478-A2.
XX
PD 14-FEB-2002.
XX
PF 06-AUG-2001; 2001WO-US24335.
XX
PR 07-AUG-2000; 2000US-223483P.
XX
PA (MONS) MONSANTO TECHNOLOGY LLC.
XX
PI Boronat A, Campos N, Rodriguez-concepcion M, Rohmer M, Seeman M;
PI Valentin HE, Venkatesh TV, Venkatramesh M;
XX
DR WPI; 2002-227151/28.
DR N-PSDB; AAD31203.
XX
PT gcpE nucleic acid which is an essential gene of the methyl-D-erythritol
PT phosphate pathway, encoding a fully defined GCPE protein which is
PT useful for increasing levels of tocopherol substrates in plants -
XX
PS Claim 1; Page 144-145; 155pp; English.
XX
CC The invention relates to gcpE nucleic acid molecule, an essential gene
CC of methyl-D-erythritol phosphate (MEP) pathway) that encodes rice,
CC Arabidopsis thaliana or Escherichia coli GCPE protein. gcpE is useful
CC for producing a transgenic plant such as Brassica campestris, B. napus,
CC canola, castor bean, coconut, cotton, crambe, linseed, maize, mustard,
CC oil palm, peanut, rapeseed, rice, safflower, sesame, soybean, sunflower,
CC or wheat with an increased isoprenoid (tocopherol) compound level. The
CC expression of GCPE protein in organisms increases the level of
CC tocopherol substrate such as isopentyl diphosphate and dimethylallyl
CC diphosphate biosynthesis. Transgenic organisms overexpressing GCPE
CC protein can nutritionally enhance food and feed sources. Overexpression
CC of GCPE protein in transgenic plant may provide tolerance to stresses
CC e.g., oxidative stress tolerance such as to oxygen or ozone, UV
CC tolerance, etc. gcpE may be used to obtain nucleic acid molecules from
CC the same species, and to obtain nucleic acid homologues. gcpE is also
CC used as or primers. The recombinant vectors are used in plant
CC transformation or transfection. gcpE an also act as markers capable of
CC detecting polymorphisms such as single nucleotide polymorphisms (SNPs).
CC gcpE is also used to determine the level or pattern of expression of
CC the protein. The present sequence is Escherichia coli GCPE protein.
XX
SQ Sequence 372 AA;

Query Match 100.0%; Score 1866; DB 23; Length 372;
Best Local Similarity 100.0%; Pred. No. 7.6e-181;
Matches 372; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MHNQAPIQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MHNQAPIQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVR 60

Qy         61 VSVPTMDAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRM 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VSVPTMDAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRM 120

Qy        121 VVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSV 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 VVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSV 180

Qy        181 KASDVFLAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 KASDVFLAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSL 240

Qy        241 AADPVEEIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AADPVEEIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDV 300
```

```
Qy        301 SIIGCVVNGPGEALVSTLGVTGGNKKSGLYEDGVRKDRLDNNDMIDQLEARIRAKASQLD 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SIIGCVVNGPGEALVSTLGVTGGNKKSGLYEDGVRKDRLDNNDMIDQLEARIRAKASQLD 360

Qy        361 EARRIDVQQVEK 372
              ||||||||||||
Db        361 EARRIDVQQVEK 372
```

RESULT 3
ABB47958
ID ABB47958 standard; Protein; 368 AA.
XX
AC ABB47958;
XX
DT 05-FEB-2002 (first entry)
XX
DE Listeria monocytogenes protein #662.
XX
KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
KW vitamin B12; bacterial infection; disease.
XX
OS Listeria monocytogenes.
XX
PN WO200177335-A2.
XX
PD 18-OCT-2001.
XX
PF 11-APR-2001; 2001WO-FR01118.
XX
PR 11-APR-2000; 2000FR-0004629.
XX
PA (INSP) INST PASTEUR.
XX
PI Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;
PI Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;
PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;
PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
PI Maduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;
PI Rose M, Voss H;
XX
DR WPI; 2002-010914/01.
XX
PT Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
PT and prevention of Listeria and related bacterial infections, and
PT related polypeptides -
XX
PS Claim 6; SEQ ID No 663; 192pp; French.
XX
CC The present invention relates to the genome sequence of Listeria
CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of
CC it are useful for selecting probes and primers for detecting genes in L.
CC monocytogenes and related organisms, and for studying genetic
CC polymorphisms and other genomes. The present sequence is a protein
CC encoded by the genome sequence of the present invention. Proteins
CC expressed from the genome sequence are useful for raising specific
CC antibodies, identification of L. monocytogenes and related organisms, and
CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin
CC B12. The genome sequence and proteins encoded by it are also useful for
CC selecting compounds that regulate gene expression and cell replication
CC and modulate L. monocytogenes-related diseases. In addition, the genome
CC sequence and proteins encoded by it are useful in pharmaceutical and
CC vaccines compositions for the treatment or prevention of infections by L.
CC monocytogenes and related organisms.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 368 AA;

Query Match 46.5%; Score 868; DB 23; Length 368;
Best Local Similarity 49.2%; Pred. No. 2.6e-79;
Matches 178; Conservative 67; Mismatches 115; Indels 2; Gaps 2;

```
Qy         10 RKSTR-IYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVRVSVPTMDA 68
              |::|| : |||: ||    : :|||| |:| |||||| :|  ||  |  ||||: |   |
Db          7 RENTRPVQVGNLTIGGSEELTIQSMTTTKTHDVEATVAEIHRLEEAGCQIVRVACPDERA 66

Qy         69 AEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRMVVDCARDK 128
              | |   ||:::::|||||||||||:|||  : ||| :||||||||  :|:  ||: |: |
Db         67 ANALSAIKKKIHIPLVADIHFDYRLALKAIDAGVDKIRINPGNIGRRDRVEKVVNAAKAK 126

Qy        129 NIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSVKASDVFLA 188
              |||||||||||||| : :||| ||   ::|||: |:  |: |:|    ||:||||| ||
Db        127 NIPIRIGVNAGSLEKKIIQKYGYPTADGMVESALAHIKILEDLDFYDIIVSLKASDVNLA 186

Qy        189 VESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSLAADPVEEI 248
              :|:|   ::  : |||||||:|   :| :||| ||| :|| |||:||||||:|||||||
Db        187 IEAYDKASRAFNYPLHLGITESGTQFAGGIKSAAGLGAILSLGIGNTLRVSLSADPVEEI 246

Qy        249 KVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDVSIIGCVVN 308
              ||  ::|||  : |    |:|||| | | |:|   | :|  :  |  |: |:::|| ||
Db        247 KVAREVLKSFGLSSNAAMLISCPTCGRIEIDLIRIANEVENYIAKIEVPIKVAVLGCAVN 306

Qy        309 GPGEALVSTLGVTGGNKKSGLYEDGVRKDRLDNNDMIDQLEARIRAKASQLDEARRIDVQ 368
              |||||  : :|: | | :  |:  |    ::    |||:|:  |   | :    ::||::
Db        307 GPGEAREADIGIAGSNGEGLLFRHGKIIRKVPEAIMIDELKKEIDILAEEF-FVKKIDLE 365

Qy        369 QV 370
              :
Db        366 SL 367
```

RESULT 4
ABP65322
ID ABP65322 standard; Protein; 403 AA.
XX
AC ABP65322;
XX
DT 19-NOV-2002 (first entry)
XX
DE Bifidobacterium longum NCC2705 ORF amino acid sequence SEQ ID NO:66.
XX
KW Bifidobacterium longum NCC2705; Bifidobacterium; bacterial;
KW antidiarrheic; antibacterial; inhibitor of Salmonella; detection;
KW identification; lactic acid bacterium; diarrhoea; pathogenic bacteria;
KW rotavirus; food composition; pharmaceutical composition.
XX
OS Bifidobacterium longum.
XX
PN EP1227152-A1.
XX
PD 31-JUL-2002.
XX
PF 30-JAN-2001; 2001EP-0102050.
XX
PR 30-JAN-2001; 2001EP-0102050.
XX
PA (NEST) SOC PROD NESTLE SA.
XX
DR WPI; 2002-668397/72.
XX
PT Novel polynucleotide comprising Bifidobacterium genome sequence useful
PT as a probe or primer for detecting and/or identifying Bifidobacterium
PT longum in a biological sample -
XX
PS Claim 3; SEQ ID 66; 80pp; English.
XX
CC The present invention describes a polynucleotide (I) comprising a
CC sequence of a Bifidobacterium genome selected from the nucleotide
CC sequences given in ABQ81842 and ABQ81843, or a sequence exhibiting at
CC least 90% identity or which hybridises with the sequences given in
CC ABQ81842 and ABQ81843. Also described is a polynucleotide (II) encoding
CC a fusion protein, comprising a sequence selected from 1097 sequences

CC given in ABP65258 to ABP66354 ligated in frame to a polynucleotide
CC encoding a heterologous polypeptide. (I) has antidiarrheic and
CC antibacterial activities, and can be used as an inhibitor of Salmonella.
CC (I) (which is a probe) is useful for the detection and/or identification
CC of Bifidobacterium longum in a biological sample. A carrier containing
CC the lactic acid bacterium Bifidobacterium longum NCC2705 (CNCM I-2618)
CC can be used for preventing and/or treating diarrhoea brought about by
CC pathogenic bacteria and/or rotavirus. The carrier is a food composition
CC selected from milk, yogurt, curd, cheese, fermented milks, milk based
CC fermented products, ice-creams, fermented cereal based products, milk
CC based powders, infant formula, pet food or a pharmaceutical composition
CC selected from tablets, liquid bacterial suspensions, dried oral
CC supplement, wet oral supplement, dry tube feeding or wet tube feeding.
CC (I) is useful in DNA arrays or chips to carry out analysis of the
CC expression of the Bifidobacterium gene. ABQ81844 to ABQ81850 represent
CC Bifidobacterium related nucleotide sequences given in the Sequence
CC Listing from the present invention but not mentioned further within the
CC specification.
CC N.B. The sequence data for this patent is not represented in the printed
CC specification but is based on sequence information supplied by the
CC European Patent Office.
XX
SQ Sequence 403 AA;

Query Match 43.6%; Score 814.5; DB 23; Length 403;
Best Local Similarity 46.3%; Pred. No. 8.4e-74;
Matches 167; Conservative 70; Mismatches 111; Indels 13; Gaps 4;

Qy 3 NQAPIQ-RRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVRV 61
Db 22 SESPLHPRRKSRRIMVGPVPVGGGAPISVQSMTNTLTANVPATLQQIAELTAAGCDIVRV 81

Qy 62 SVPTMDAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRM- 120
Db 82 AVPSQDDADALPEICRKSPIPVIADIHFQSKYVFQAIDAGCAAVRVNPGNIRKFDEVGPD 141

Qy 121 VVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSV 180
Db 142 ICKAATDAGISLRIGVNAGSLDKELYAKYGGPTPEALVASALKEAHMFEDVGFHDFKISV 201

Qy 181 KASDVFLAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSL 240
Db 202 KHHDVITMVETYRLLASKGDWPLHLGVTEAGPAWQGTIKSCLAFGALLAEGIGDTIRVSL 261

Qy 241 AADPVEEIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDV 300
Db 262 SAPPAEEVKVGCKLLEYMGLRPRKFDIISCPSCGRAQVDVIQLASAVTEGLKDVTAPIRV 321

Qy 301 SIIGCVVNGPGEALVSTLGVTGGNKKSGLYEDG-----VRKDRLDN------NDMIDQLE 349
Db 322 AVMGCIVNGPGEAREADLGVASGNGKGQIFIKGKVIKTVPEDQIVDTLLTIANDIAAQME 381

Qy 350 A 350
Db 382 A 382

RESULT 5
AAW98298
ID AAW98298 standard; Protein; 359 AA.
XX
AC AAW98298;
XX
DT 31-MAR-1999 (first entry)
XX
DE H. pylori GHPO 76 protein.
XX
KW GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
KW peptic ulcer disease.
XX
OS Helicobacter pylori.
XX
PN WO9843478-A1.
XX
PD 08-OCT-1998.
XX
PF 01-APR-1998; 98WO-US06371.
XX
PR 29-JUL-1997; 97US-0902615.
PR 01-APR-1997; 97US-0833457.
PR 24-JUN-1997; 97US-0881227.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
XX
PI Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;
XX
DR WPI; 1998-542293/46.
DR N-PSDB; AAX14017.
XX
PT New isolated Helicobacter polynucleotides - used to develop products
PT for the diagnosis, prevention and treatment of Helicobacter
PT infections and gastrointestinal diseases
XX
PS Claim 8; Page 372-374; 2054pp; English.
XX
CC This sequence represents a Helicobacter pylori GHPO protein of the
CC invention. The polypeptides can be used for preventing or treating
CC Helicobacter infections, and gastroduodenal diseases associated with
CC these infections, including acute, chronic, and atrophic gastritis, and
CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
CC used for the production of antibodies. The products can also be used for
CC detection and diagnosis.
XX
SQ Sequence 359 AA;

Query Match 42.6%; Score 794; DB 19; Length 359;
Best Local Similarity 45.8%; Pred. No. 8.5e-72;
Matches 162; Conservative 73; Mismatches 117; Indels 2; Gaps 2;

Qy 9 RRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVRVSVPTMDA 68
Db 5 RVKTKQIFIGGVAIGGDAPISTQSMTFSKTADIESTKNQIDRLKLAGADLVRVAVSNEKD 64

Qy 69 AEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRMVVDCARDK 128
Db 65 ALALKELKKVSPLPLIADIHFHYKFAL-IAAQSVDAIRINPGNIGSKEKIKAVVDACKEK 123

Qy 129 NIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSVKASDVFLA 188
Db 124 NIPIRIGVNAGSLEKQFDQKYG-PTPKGMVESALYNAKLLEDLDFTNFKISLKASDVIRT 182

Qy 189 VESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSLAADPVEEI 248
Db 183 IEAYRMLRPLVIYPFHLGVTEAGNLFSSSIKSAMALGGLLMEGIGDTMRVSITGELENEI 242

Qy 249 KVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDVSIIGCVVN 308
Db 243 KVARAILRHSGRLKEGINWISCPTCGRIEANLVDMAIKVEKRLSHIKTPLDISVMGCVVN 302

Qy 309 GPGEALVSTLGVTGGNKKSGLYEDGVRKDRLDNNDMIDQLEARIRAKASQLDEA 362
Db 303 ALGEAKHADMAIAFGNRSGLIIKEGKVIHKLAEKDLFETFVIEVENLAKEREKS 356

RESULT 6
AAU56212
ID AAU56212 standard; Protein; 415 AA.
XX
AC AAU56212;
XX
DT 27-FEB-2002 (first entry)
XX
DE Propionibacterium acnes immunogenic protein #17108.
XX
KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;

KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.
XX
OS Propionibacterium acnes.
XX
PN WO200181581-A2.
XX
PD 01-NOV-2001.
XX
PF 20-APR-2001; 2001WO-US12865.
XX
PR 21-APR-2000; 2000US-199047P.
PR 02-JUN-2000; 2000US-208841P.
PR 07-JUL-2000; 2000US-216747P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
DR WPI; 2001-616774/71.
DR N-PSDB; AAS59573.
XX
PT Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris -
XX
PS Example 1; SEQ ID No 17407; 1069pp; English.
XX
CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA).
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 415 AA;

Query Match 42.1%; Score 785.5; DB 22; Length 415;
Best Local Similarity 44.2%; Pred. No. 7.9e-71;
Matches 159; Conservative 68; Mismatches 132; Indels 1; Gaps 1;

```
Qy          9 RRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVRVSVPTMDA 68
              |||: :| ||:| :|  |||:||||| |:| || ||: || ||   | ||||: ||
Db         53 RRKTHQIKVGDVLVGGDAPISVQSMTTTKTHDVGATLQQIAALTAAGCDIVRVACPTDKD 112

Qy         69 AEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGN-EERIRMVVDCARD 127
              ||   :| ::  :|::|||||  :   :  | |   :|:|||||   :::|  :   | :
Db        113 AEVLPIIAKRSQIPVIADIHFQPKYVFQAIEAGCGAVRVNPGNIRKFDDQIESICQAATE 172

Qy        128 KNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSVKASDVFL 187
              ||||||||||:| | :||| || :|::|||:      ::  |  ||:|||  |  :
Db        173 HGTSIRIGVNAGSLDKRLLDKYGAPTAEAMVESALWEASLFEQYGFRDFKISVKHHDPVV 232

Qy        188 AVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSLAADPVEE 247
              : :|  || : | |||||:|||| |  | :|||:  | ||:||||||:||||:||||||
Db        233 MIRAYEQLAAKCDYPLHLGVTEAGPAFQGTIKSAVAFGHLLAEGIGDTIRVSLSADPVEE 292

Qy        248 IKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDVSIIGCVV 307
              :|||  ||:|| :| ||:  ::||:| | : ||:   | :   || |  |: |:::||||
Db        293 VKVGIKILESLNLRPRGLEIVSCPSCGRCQVDVLTLANDVTDALEGIDAPLRVAVMGCVV 352

Qy        308 NGPGEALVSTLGVTGGNKKSGLYEDGVRKDRLDNNDMIDQLEARIRAKASQLDEARRIDV 367
              || ||   : |||  || |  ::: |     :   :::  |       | ::|    ::|
Db        353 NGLGEGREADLGVAAGNGKGKIFKHGEVIRTVPEGEIVQFLVQEANRMADEMDTTGAVEV 412
```

RESULT 7
AAG91951
ID AAG91951 standard; Protein; 378 AA.
XX
AC AAG91951;
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum protein fragment SEQ ID NO: 5705.
XX
KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis.
XX
OS Corynebacterium glutamicum.
XX
PN EP1108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-0127688.
XX
PR 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
DR WPI; 2001-376931/40.
DR N-PSDB; AAH67170.
XX
PT Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT -
XX
PS Claim 17; SEQ ID NO: 5705; 246pp + Sequence Listing; English.
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 378 AA;

Query Match 40.8%; Score 761; DB 22; Length 378;
Best Local Similarity 44.3%; Pred. No. 2.1e-68;
Matches 159; Conservative 64; Mismatches 124; Indels 12; Gaps 3;

```
Qy          9 RRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVRVSVPTMDA 68
              |||: :: || | :|   ||:||||| |:| |:  |: ||  |.   | |||||: |
Db          5 RRKTRQLMVGKVGVGSDHPISVQSMTTTKTHDINGTLQQIAQLTATGCDIVRVACPKTVD 64
```

```
Qy         69 AEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGN-EERIRMVVDCARD 127
Db         65 AEALPIIAKKSPIPVIADIHFQPKYIFAAIDAGCAAVRVNPGNIKEFDGRVKEVAKAAGD 124

Qy        128 KNIPIRIGVNAGSLEKDLQEKY-GEPTPQALLESAMRHVDHLDRLNFDQFKVSVKASDVF 186
Db        125 AGIPIRIGVNGGSLDKRTLDKYHGKATPEALVESAMWEAGLFEEHGFGDIAISVKHSDPV 184

Qy        187 LAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSLAADPVE 246
Db        185 LMVEAYRQLAEQSDYPLHLGVTEAGPKFMGTIKSSVAFGALLSQGIGDTIRVSLSADPVE 244

Qy        247 EIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDVSIIGCV 306
Db        245 EIKVGDQILQSLNLRPRKLEIVSCPSCGRAQVDVYSLAEEVTEALDGMEVPLRVAVMGCV 304

Qy        307 VNGPGEALVSTLGVTGGNKKSGLYEDGVRKDRLDNNDMIDQLEARIRAKASQLDEARRI 365
Db        305 VNGPGEARDADLGVASGNGKGQIFVKGEVIKTVPESQIVETL----------IEEAMRI 353
```

RESULT 8
AAG81205
ID AAG81205 standard; Protein; 387 AA.
XX
AC AAG81205;
XX
DT 04-SEP-2001 (first entry)
XX
DE Mycobacterium tuberculosis potential drug target protein SEQ ID 256.
XX
KW Drug target; growth; organism viability; characterisation.
XX
OS Mycobacterium tuberculosis.
XX
PN WO200135317-A1.
XX
PD 17-MAY-2001.
XX
PF 13-NOV-2000; 2000WO-US31152.
XX
PR 12-NOV-1999; 99US-0165086.
PR 12-NOV-1999; 99US-0165124.
PR 01-FEB-2000; 2000US-0179531.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Eisenberg D, Rotstein SH, Marcotte EM;
XX
DR WPI; 2001-329193/34.
DR N-PSDB; AAH52056.
XX
PT Identifying nucleotide or polypeptide sequence for use as drug target,
PT involves providing algorithm that analyzes a functional relationship
PT between nucleotide or polypeptide sequences, and comparing the
PT sequences -
XX
PS Disclosure; Page 182; 207pp; English.
XX
CC This invention relates to a method for identifying a nucleotide or
CC polypeptide sequence that may be a drug target, or essential for growth
CC or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092
CC represent DNA encoding proteins AAG81096 - AAG81241, Mycobacterium
CC tuberculosis proteins which are potential drug targets. The DNA and
CC protein sequences are used to illustrate the method of the invention. The
CC method involves providing an unknown nucleotide or polypeptide sequences,
CC and comparing it to a number of sequences along with at least one
CC algorithm capable of analysing a functional relationship between
CC nucleotide and polypeptide sequences. The method is useful for
CC characterising the function of nucleic acids and polypeptides that may be
CC useful as a target for a drug or essential for the growth or viability of
CC an organism.
XX

SQ Sequence 387 AA;

Query Match 40.3%; Score 751.5; DB 22; Length 387;
Best Local Similarity 42.7%; Pred. No. 2e-67;
Matches 156; Conservative 64; Mismatches 134; Indels 11; Gaps 2;

```
Qy          9 RRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVRVSVPTMDA 68
Db         18 RRATRQLMVGNVGVGSDHPVSVQSMCTTKTHDVNSTLQQIAELTAAGCDIVRVACPRQED 77

Qy         69 AEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGN-EERIRMVVDCARD 127
Db         78 ADALAEIARHSQIPVVADIHFQPRYIFAAIDAGCAAVRVNPGNIKEFDGRVGEVAKAAGA 137

Qy        128 KNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSVKASDVFL 187
Db        138 AGIPIRIGVNAGSLDKRFMEKYGKATPEALVESALWEASLFEEHGFGDIKISVKHNDPVV 197

Qy        188 AVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSLAADPVEE 247
Db        198 MVAAYELLAARCDYPLHLGVTEAGPAFQGTIKSAVAFGALLSRGIGDTIRVSLSAPPVEE 257

Qy        248 IKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDVSIIGCVV 307
Db        258 VKVGNQVLESLNLRPRSLEIVSCPSCGRAQVDVYTLANEVTAGLDGLDVPLRVAVMGCVV 317

Qy        308 NGPGEALVSTLGVTGGNKKSGLYEDGVRKDRLDNNDMIDQLEARIRAKASQLDEARRIDV 367
Db        318 NGPGEAREADLGVASGNGKGQIFVRGEVIKTVPEAQIVETL----------IEEAMRLAA 367

Qy        368 QQVEK 372
Db        368 EMGEQ 372
```

RESULT 9
ABP57500
ID ABP57500 standard; Protein; 387 AA.
XX
AC ABP57500;
XX
DT 28-APR-2003 (first entry)
XX
DE Mycobacterium tuberculosis protein SEQ ID NO:129.
XX
KW Mycobacterium tuberculosis; infection; antibacterial; tuberculostatic;
KW immunostimulant; vaccine; gene therapy; mycobacterial infection.
XX
OS Mycobacterium tuberculosis.
XX
PN WO2003000721-A2.
XX
PD 03-JAN-2003.
XX
PF 21-JUN-2002; 2002WO-GB02845.
XX
PR 22-JUN-2001; 2001GB-0015365.
PR 07-SEP-2001; 2001GB-0021780.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI James BW, Bacon J, Marsh P;
XX
DR WPI; 2003-201403/19.
DR N-PSDB; ABZ71126.
XX
PT New mycobacterial peptide, its fragment, variant or derivative, useful
PT as vaccine for treating or preventing mycobacterial infections, and as
PT diagnostic reagents for identifying such infections -
XX
PS Claim 2; Page 235-236; 246pp; English.
XX
CC ABP57436 to ABP57504 represent mycobacterial amino acid sequences (I)

```
CC   encoded by ABZ71062 to ABZ71130 (II), which are isolated from
CC   Mycobacterium tuberculosis. (I) are encoded by genes (II) whose
CC   expression is induced or up-regulated during culture of a mycobacterium
CC   under conditions defined by a dissolved oxygen tension of at least 10%
CC   air saturation measured at 37 plus degrees Celsius, when compared with a
CC   dissolved oxygen tension of at least 40% air saturation measured at 37
CC   plus degrees Celsius. (I) and (II) have antibacterial, tuberculostatic
CC   and immunostimulant activities, and can be used in vaccines and gene
CC   therapy. (I) and (II) can be used for the manufacture of a medicament
CC   for treating or preventing a mycobacterial infection. They can also be
CC   used for the manufacture of a diagnostic reagent for identifying a
CC   mycobacterial infection.
XX
SQ   Sequence   387 AA;

  Query Match             40.3%;  Score 751.5;  DB 24;  Length 387;
  Best Local Similarity   42.7%;  Pred. No. 2e-67;
  Matches  156;  Conservative   64;  Mismatches  134;  Indels   11;  Gaps    2;

Qy          9 RRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVRVSVPTMDA 68
              || : :: |||| :|   |::|||| |:| || :|: ||  |   | |||||: |  :
Db         18 RRATRQLMVGNVGVGSDHPVSVQSMCTTKTHDVNSTLQQIAELTAAGCDIVRVACPRQED 77

Qy         69 AEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGN-EERIRMVVDCARD 127
              |:|   | :   :|:||||||  |      : |   :|:|||||   : |:  |   |
Db         78 ADALAEIARHSQIPVVADIHFQPRYIFAAIDAGCAAVRVNPGNIKEFDGRVGEVAKAAGA 137

Qy        128 KNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSVKASDVFL 187
              ||||||||||||:|   ||||: ||:||:|||:      :   |   |:||| :|  :
Db        138 AGIPIRIGVNAGSLDKRFMEKYGKATPEALVESALWEASLFEEHGFGDIKISVKHNDPVV 197

Qy        188 AVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSLAADPVEE 247
              | :| ||| : | |||||:|||| |  | :|||:  | ||| |||||:||||:| ||||
Db        198 MVAAYELLAARCDYPLHLGVTEAGPAFQGTIKSAVAFGALLSRGIGDTIRVSLSAPPVEE 257

Qy        248 IKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDVSIIGCVV 307
              :|||  :|:|| :| | :  ::||:| | : ||    | :   |: :  |: |:::||||
Db        258 VKVGNQVLESLNLRPRSLEIVSCPSCGRAQVDVYTLANEVTAGLDGLDVPLRVAVMGCVV 317

Qy        308 NGPGEALVSTLGVTGGNKKSGLYEDGVRKDRLDNNDMIDQLEARIRAKASQLDEARRIDV 367
              ||||||  : |||  || |  ::  |    :   ::: |          ::|| |:
Db        318 NGPGEAREADLGVASGNGKGQIFVRGEVIKTVPEAQIVETL----------IEEAMRLAA 367

Qy        368 QQVEK 372
              :  |:
Db        368 EMGEQ 372


RESULT 10
ABP78786
ID   ABP78786 standard; Protein; 421 AA.
XX
AC   ABP78786;
XX
DT   07-MAR-2003  (first entry)
XX
DE   N. gonorrhoeae amino acid sequence SEQ ID 4102.
XX
KW   Antibacterial; infection; vaccine; gene therapy.
XX
OS   Neisseria gonorrhoeae.
XX
PN   WO200279243-A2.
XX
PD   10-OCT-2002.
XX
PF   12-FEB-2002; 2002WO-IB02069.
XX
PR   12-FEB-2001; 2001GB-0003424.
XX
PA   (CHIR-) CHIRON SPA.
XX
PI   Fontana MR,  Pizza M,  Masignani V,  Monaci E;
XX
DR   WPI; 2003-058415/05.
DR   N-PSDB; ABZ39756.
XX
PT   New protein from Neisseria gonorrheae, useful for the manufacture of a
PT   medicament for treating or preventing N. gonorrheae infection  -
XX
PS   Disclosure; Page 487; 815pp; English.
XX
CC   The present invention relates to proteins from Neisseria gonorrhoeae.
CC   Also disclosed are the nucleic acid molecules encoding the proteins and
CC   antibodies that specifically bind to the proteins. The composition
CC   comprising the protein, nucleic acid or antibody is useful for the
CC   manufacture of a medicament for treating or preventing N. gonorrhoeae
CC   infection, this may be in the form of a vaccine or gene therapy.
CC   Sequences given in records ABP76736-ABP81046 represent nucleic acid
CC   molecules of the invention.
XX
SQ   Sequence   421 AA;

  Query Match             28.6%;  Score 534;  DB 24;  Length 421;
  Best Local Similarity   32.3%;  Pred. No. 3.1e-45;
  Matches  135;  Conservative   87;  Mismatches  142;  Indels   54;  Gaps   12;

Qy          7 IQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVRVSVPTM 66
              :||||: :: : :: :|  ||: :|||||| | | :||  ||| |   |:::||::| .:
Db          4 LQRRKTHQVLIDHITVGSEAPVVIQSMTNTDTADAKATALQIKELSDAGSEMVRITVNSP 63

Qy         67 DAAEAFKLIKQQVN-----VPLVADIHFD-YRIALKVAEYG--VDCLRINPGNIG----N 114
              :||     |:::::      ||: | ||:  |:  :  | |  :   ||||||:|
Db         64 EAASKVAEIRRRLDDMGYATPLIGDFHFNGERLLAEFPECGKALSKYRINPGNVGKGVKG 123

Qy        115 EERIRMVVDCARDKNIPIRIGVNAGSLEKDLQEKY-------GEPTP------QALLESA 161
              :|:   ::  | : :  :|||||| |||:: | ::           | |      :||: ||
Db        124 DEKFAFMIRTAAENDKAVRIGVNWGSLDQSLAKRMMDANLVSSAPKPPEEVMKEALIVSA 183

Qy        162 MRHVDHLDRLNFDQFKV--SVKASDVFLAVESYRLLAKQIDQPLHLGITEAGGARSGAVK 219
              :   :    |   : |:  | | | |   :: || |  :   |||||:||||    | |
Db        184 LESAEKAVLLGLPEDKIILSCKVSAVHDLIQVYRELGSRCVYPLHLGLTEAGMGSKGIVA 243

Qy        220 SAIGLGLLLSEGIGDTLRVSLAADP----VEEIKVGFDILKSLRIRSRGINFIACPTCSR 275
              |   | :|| ||||||:|:||  :|     :|: || :||::: :||       ||| | |
Db        244 STAALSVLLQEGIGDTIRISLTPEPGSPRTQEVVVGQEILQTMGLRSFTPMVTACPGCGR 303

Qy        276 QEFDVIGTV-----NALEQR------LEDIITPMDVSIIGCVVNGPGEALVSTLGV---- 320
                  |   :     | | |:      |   :  ::|:::||||||||||: :: :|:
Db        304 TTSTVFQELAQDVQNYLRQKMSIWRTLYPGVESLNVAVMGCVVNGPGESKLADIGISLPG 363

Qy        321 TGGNKKSGLYEDGVRKDRLDNNDMIDQLEARIR--------AKASQLDEARRIDVQQV 370
              ||    : :| || ||  |  |::  :  | :           :|: :: : | :| :
Db        364 TGETPVAPVYVDGERKVTLKGNNIASEFLAIVEEYVKTNYGKNSSKRNKGKVIPIQSL 421


RESULT 11
AAY37164
ID   AAY37164 standard; Protein; 611 AA.
XX
AC   AAY37164;
XX
DT   07-OCT-1999  (first entry)
XX
DE   Amino acid sequence of a Chlamydia trachomatis protein.
XX
KW   Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
KW   paratrachoma; inclusion conjunctivitis; genital disease; perihepatitis;
KW   nongonococcal uretritis; epidymitis; cervicitis; salpingitis;
KW   bartholinitis; pneumopathy; venereal lymphogranulomatosis.
XX
OS   Chlamydia trachomatis.
XX
PN   WO9928475-A2.
XX
PD   10-JUN-1999.
```

```
XX
PF   27-NOV-1998;   98WO-IB01939.
XX
PR   04-NOV-1998;   98US-0107077.
PR   28-NOV-1997;   97FR-0015041.
PR   17-DEC-1997;   97FR-0016034.
XX
PA   (GEST ) GENSET.
XX
PI   Griffais R;
XX
DR   WPI; 1999-371125/31.
XX
PT   Genome sequence of Chlamydia trachomatis
XX
PS   Disclosure; Page 943-944; 1755pp; English.
XX
CC   AAY36754-Y37949 are encoded by open reading frames (ORFs) of the genome
CC   of Chlamydia trachomatis (see AAZ01425). The polypeptides can be used as
CC   vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences
CC   can also be used to control growth of the microorganism. Chlamydia
CC   trachomatis is responsible for a large number of diseases, e.g. eye
CC   diseases such as conventional trachoma, nonendemic trachoma,
CC   paratrachoma, and inclusion conjunctivitis; genital diseases such as
CC   nongonococcal uretritis, epidymitis, cervicitis, salpingitis,
CC   perihepatitis, bartholinitis; pneumopathy in breast feeding infants;
CC   and venereal lymphogranulomatosis. The polypeptides of the invention
CC   may be of use in treating these diseases.
XX
SQ   Sequence   611 AA;

  Query Match             26.3%;  Score 490;  DB 20;  Length 611;
  Best Local Similarity   23.9%;  Pred. No. 1.7e-40;
  Matches  134;  Conservative   76;  Mismatches  109;  Indels  242;  Gaps    8;

Qy          9 RRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVRVSVPTMDA 68
              |||:  : :|:: :|    | :|||| | ||||: || || ||: :| |||||:|  :
Db         20 RRKTLPVRIGDLFVGSEHSIKIQSMTTTATTDVDGTVRQICALQELGCDIVRVTVQGLRE 79

Qy         69 AEAF-----KLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGN----------I 112
                |      :||:| :::|||||||| : |: | :  || :||||||          |
Db         80 VHACEHIKDRLIQQNISIPLVADIHFFPQAAIHVVDC-VDKVRINPGNYVDKRNMFTGKI 138

Qy        113 GNEERIRMVVDCARDKNIPI----------RIGVNAGSLEKDLQEKYGEPTPQALLESA 161
               ::|:    ::   :|  |:          ||||| ||| : : ::||  | : :: ||
Db        139 YSDEQYAHSLEHLMNKFSPLVEKCKRLGKAMRIGVNHGSLSERVTQRYGN-TIEGMVYSA 197

Qy        162 MRHVDHLDRLNFDQFKVSVKASDVFLAVESYRLLAKQIDQ-----PLHLGITEAGGARSG 216
              : : :    :::     |:|:|:  : | :|| || ::||     |||||:||||   :|
Db        198 LEYAEVCVAMDYHDVIFSMKSSNPKVMVAAYRSLAYELDQREWSYPLHLGVTEAGSGTAG 257

Qy        217 AVKSAIGLGLLLSEGIGDTLRVSLAADPVEEIKV---------------------------- 250
               ||||:|:| |||||:|||:| ||   |: || :
Db        258 IVKSAVGIGTLLSEGLGDTIRCSLTGSPINEIPICIDLLKQTTELSERWGEADNPFAIHS 317

Qy        251 ------------------------------------------------------------ 250

Db        318 SKQLGTRNTLNTPPWDNVYGLLINLTDVQLLTAEPIELLQCLGIDTTTGKIDPTTPEGVV 377

Qy        251 ------------------------------------------------------------ 250

Db        378 VPKAMRSSPIVSEIEKHLLVFNKEDAPILNPMNEEEWLSEETLSAPFVYFEVTDIHTARR 437

Qy        251 ------------------------------------------------------------ 250

Db        438 FFSLRQHSTQPVCLSFSLDPHLSKNEAIIDLSARLGALLLDGLGSCVLLDFVDIKLSRTL 497

Qy        251 GFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDVSIIGCVVNGP 310
              || ||:|  |||  : :::|| | |  ||::     : :| : :   : ::::||:|||
Db        498 GFLILQSANIRSVTVEYVSCPGCGRTLFDLLAVSQRIRERTKHLPGGLKIAVMGCIVNGP 557

Qy        311 GEALVSTLGVTGGNKKSGLYE 331
              ||   :  |  |   | |: :
```

```
Db        558 GEMADADFGYVG--SKPGMID 576


RESULT 12
AAY34971
ID   AAY34971 standard; Protein; 621 AA.
XX
AC   AAY34971;
XX
DT   13-SEP-1999  (first entry)
XX
DE   Amino acid sequence of a Chlamydia pneumoniae  protein.
XX
KW   Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KW   sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KW   vaccine; neutralising epitope.
XX
OS   Chlamydia pneumoniae.
XX
PN   WO9927105-A2.
XX
PD   03-JUN-1999.
XX
PF   20-NOV-1998;   98WO-IB01890.
XX
PR   04-NOV-1998;   98US-0107078.
PR   21-NOV-1997;   97FR-0014673.
XX
PA   (GEST ) GENSET.
XX
PI   Griffais R;
XX
DR   WPI; 1999-357842/30.
XX
PT   Genome sequence of Chlamydia pneumoniae
XX
PS   Page 885-886; Disclosure; 1912pp; English.
XX
CC   AAY34584-Y35879 represent the proteins encoded by all the open reading
CC   frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.
CC   C. pneumoniae causes respiratory disease such as pneumonia and
CC   bronchitis and is thought to be a contributing factor in heart
CC   disease, sarcoidosis, sinusitis, purulent otitis media, erythema
CC   nodosum or pharyngitis. The polypeptides encoded by the open reading
CC   frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in
CC   immunogenic compositions as vaccines. Vectors containing C. pneumoniae
CC   nucleotides sequences can also be used as immunogenic compositions,
CC   especially where the vector directs the expression of a neutralising
CC   epitope of C. pneumoniae.
XX
SQ   Sequence   621 AA;

  Query Match             25.2%;  Score 470.5;  DB 20;  Length 621;
  Best Local Similarity   38.9%;  Pred. No. 1.7e-38;
  Matches  111;  Conservative   55;  Mismatches   84;  Indels   35;  Gaps    5;

Qy          9 RRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVRVSVPTMDA 68
              |||:  : :|:: ||    |  |||| | |||:::|| || ||     |||||:|  :
Db         20 RRKTHTVRIGNLYIGSDHSIKTQSMTTTLTTDIDSTVEQIYALAEHNCDIVRVTVQGIKE 79

Qy         69 AEAFKLIKQQ-----VNVPLVADIHFDYRIALKVAEYGVDCLRINPGN---------- 111
              |:| : ||::     :|:|||||||  : |: ||::  | :||||||
Db         80 AQACEKIKERLIALGLNIPLVADIHFFPQAAMLVADF-ADKVRINPGNYIDKRNMFKGTK 138

Qy        112 -----------IGNEERIRMVVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLES 160
                        :  ||:   :|:  :      :|||||  ||| : : :|||: | : :: |
Db        139 IYTEASYAQSLLRLEEKFAPLVEKCKRLGKAMRIGVNHGSLSERIMQKYGD-TIEGMVAS 197

Qy        161 AMRHVDHLDRLNFDQFKVSVKASDVFLAVESYRLLAKQID-----QPLHLGITEAGGARS 215
              |: ::   ::|||:     |:|:|:  : | :|| ||| :|      |||||:||||
Db        198 AIEYIAVCEKLNYRDVVFSMKSSNPKIMVTAYRQLAKDLDARGWLYPLHLGVTEAGMGVD 257

Qy        216 GAVKSAIGLGLLLSEGIGDTLRVSLAADPVEEIKVGFDILKSLRI 260
```

```
                | :|||:|:| ||:||:|||:| ||   |  || |   :|:  :|
Db        258 GIIKSAVGIGTLLAEGLGDTIRCSLTGCPTTEIPVCDSLLRHTKI 302


RESULT 13
AAE19650
ID   AAE19650 standard; Protein; 740 AA.
XX
AC   AAE19650;
XX
DT   31-MAY-2002  (first entry)
XX
DE   Arabidopsis thaliana GCPE protein.
XX
KW   gcpE gene; methyl-D-erythritol phosphate; MEP; dimethylallyl diphosphate;
KW   transgenic plant; isoprenoid compound; tocopherol; isopentyl diphosphate;
KW   food; feed source; transfection; single nucleotide polymorphism; SNP;
KW   oxidative stress tolerance; UV tolerance; transformation; chromosome V;
KW   plant; GCPE protein.
XX
OS   Arabidopsis thaliana.
XX
FH   Key             Location/Qualifiers
FT   Cleavage-site   32..33
FT                   /note= "Plastidial targeting peptide cleavage site"
XX
PN   WO200212478-A2.
XX
PD   14-FEB-2002.
XX
PF   06-AUG-2001; 2001WO-US24335.
XX
PR   07-AUG-2000; 2000US-223483P.
XX
PA   (MONS ) MONSANTO TECHNOLOGY LLC.
XX
PI   Boronat A,  Campos N,  Rodriguez-concepcion M,  Rohmer M,  Seeman M;
PI   Valentin HE,  Venkatesh TV,  Venkatramesh M;
XX
DR   WPI; 2002-227151/28.
DR   N-PSDB; AAD31201.
XX
PT   gcpE nucleic acid which is an essential gene of the methyl-D-erythritol
PT   phosphate pathway, encoding a fully defined GCPE protein which is
PT   useful for increasing levels of tocopherol substrates in plants  -
XX
PS   Claim 1; Page 140-142; 155pp; English.
XX
CC   The invention relates to gcpE nucleic acid molecule, an essential gene
CC   of methyl-D-erythritol phosphate (MEP) pathway) that encodes rice,
CC   Arabidopsis thaliana or Escherichia coli GCPE protein. gcpE is useful
CC   for producing a transgenic plant such as Brassica campestris, B. napus,
CC   canola, castor bean, coconut, cotton, crambe, linseed, maize, mustard,
CC   oil palm, peanut, rapeseed, rice, safflower, sesame, soybean, sunflower,
CC   or wheat with an increased isoprenoid (tocopherol) compound level. The
CC   expression of GCPE protein in organisms increases the level of
CC   tocopherol substrate such as isopentyl diphosphate and dimethylallyl
CC   diphosphate biosynthesis. Transgenic organisms overexpressing GCPE
CC   protein can nutritionally enhance food and feed sources. Overexpression
CC   of GCPE protein in transgenic plant may provide tolerance to stresses
CC   e.g., oxidative stress tolerance such as to oxygen or ozone, UV
CC   tolerance, etc. gcpE may be used to obtain nucleic acid molecules from
CC   the same species, and to obtain nucleic acid homologues. gcpE is also
CC   used as or primers. The recombinant vectors are used in plant
CC   transformation or transfection. gcpE an also act as markers capable of
CC   detecting polymorphisms such as single nucleotide polymorphisms (SNPs).
CC   gcpE is also used to determine the level or pattern of expression of
CC   the protein. The present sequence is Arabidopsis thaliana GCPE protein.
CC   gcpE gene is located on chromosome V.
XX
SQ   Sequence   740 AA;

  Query Match           24.8%;  Score 462;  DB 23;  Length 740;
  Best Local Similarity   38.6%;  Pred. No. 1.6e-37;
  Matches  105;  Conservative   47;  Mismatches   86;  Indels   34;  Gaps    5;

Qy          9 RRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVRVSVPTMDA 68
              |||:  : |||| :|   || :|:|| : | |:  ||:::  :   ||||||::|
Db         84 RRKTRTVMVGNVALGSEHPIRIQTMTTSDTKDITGTVDEVMRIADKGADIVRITVQGKKE 143

Qy         69 AEAF-----KLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRMVVD 123
              |:|      ||::   |:||||||||   :||:|||   | :|:|||||  :       :|
Db        144 ADACFEIKDKLVQLNYNIPLVADIHFAPTVALRVAEC-FDKIRVNPGNFADRRAQFETID 202

Qy        124 CARDK---------------------NIPIRIGVNAGSLEKDLQEKYGEPTPQALLESA 161
                 |:                          :||| | |||   :   ||: :|: ::|||
Db        203 YTEDEYQKELQHIEQVFTPLVEKCKKYGRAMRIGTNHGSLSDRIMSYYGD-SPRGMVESA 261

Qy        162 MRHVDHLDRLNFDQFKVSVKASDVFLAVESYRLLAKQI-----DQPLHLGITEAGGARSG 216
                  :|::  |  |:|||:  : |::||||  ::     | |||||:||||    |
Db        262 FEFARICRKLDYHNFVFSMKASNPVIMVQAYRLLVAEMYVHGWDYPLHLGVTEAGEGEDG 321

Qy        217 AVKSAIGLGLLLSEGIGDTLRVSLAADPVEEI 248
              :|||||:| || :|:|||:||||   | |||
Db        322 RMKSAIGIGTLLQDGLGDTIRVSLTEPPEEEI 353


RESULT 14
AAB45693
ID   AAB45693 standard; Protein; 679 AA.
XX
AC   AAB45693;
XX
DT   15-MAR-2001  (first entry)
XX
DE   P. falciparum gcpE protein.
XX
KW   Isoprenoid; gcpE; yfgB; antimicrobial; transgenic plant; agriculture;
KW   antimycotic; antiparasitic; antiviral; fungicidal; herbicidal.
XX
OS   Plasmodium falciparum.
XX
PN   WO200072022-A1.
XX
PD   30-NOV-2000.
XX
PF   20-MAY-2000; 2000WO-EP04592.
XX
PR   21-MAY-1999;   99DE-1023567.
PR   21-MAY-1999;   99DE-1023568.
XX
PA   (JOMA/) JOMAA H.
XX
PI   Jomaa H;
XX
DR   WPI; 2001-025196/03.
DR   N-PSDB; AAC82654.
XX
PT   Incorporating gcpE and yfgB genes into viruses and cells, for
PT   increasing isoprenoid content and identifying e.g. antimicrobial
PT   agents, comprises using DNA sequences from bacteria or parasites  -
XX
PS   Disclosure; Page 21-23; 36pp; German.
XX
CC   This invention describes a novel method for incorporating gcpE and yfgB
CC   genes into viruses and cells for increasing isoprenoid content and
CC   identifying e.g. antimicrobial agents, comprises using DNA sequences (I)
CC   from the gcpE or yfgB genes of bacteria or parasites or DNA sequences
CC   (II) which hybridize to the specified genes or encode a plastid protein
CC   with the same biological activity as those encoded by the genes. The
CC   invention also describes (1) plant cells containing (I) or (II); (2)
CC   transformed plant cells, and transgenic plants regenerated from them,
CC   that contain (I) or (II); (3) determining the enzymatic activity of a
CC   gcpE protein; or (4) screening compounds (A) that have antimycotic,
CC   antiparasitic or antiviral activity in humans or animals or antiviral,
CC   antiparasitic, fungicidal or herbicidal activity in plants. (I) and (II)
```

```
CC   are used: (i) to increase the isoprenoid levels in viruses and cells;
CC   (ii) for determining the enzymatic activity of gcpE and yfgB proteins;
CC   and (iii) to identify compounds that inhibit activity of gcpE, i.e.
CC   potential antibacterial, antimycotic, antiparasitic or antiviral agents
CC   for use in humans or animals, or antiviral, antiparasitic, fungicidal
CC   or herbicidal agents for agriculture.
XX
SQ   Sequence   679 AA;

  Query Match             24.7%;  Score 460;  DB 22;  Length 679;
  Best Local Similarity   38.9%;  Pred. No. 2.3e-37;
  Matches  112;  Conservative   57;  Mismatches   81;  Indels   38;  Gaps    8;

Qy          9 RRKSTR-IYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVRVSVPTMD 67
              :|  || : :||| ||    ||:|:| :  | :||  | ||:  : :|||||::|  :
Db        117 KRLPTREVVIGNVKIGGNNKIAIQTMASCDTRNVEECVYQIRKCKDLGADIVRLTVQGVQ 176

Qy         68 AAEAF-----KLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGN-------- 114
               |:|      ||: : ||:||| |||| :|||  |:   : :|:|||| :
Db        177 EAQASYHIKEKLLSENVNIPLVTDIHFNPKIALMAADV-FEKIRVNPGNYVDGRKKWIDK 235

Qy        115 ----------------EERIRMVVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALL 158
                             ||: : ::  | |  |  |||| | | |   :   ||: || ||:
Db        236 VYKTKEEFDEGKLFIKEKFVPLIEKCKR-LNRAIRIGTNHGFLSSRVLSYYGD-TPLALV 293

Qy        159 ESAMRHVDHLDRLNFDQFKVSVKASDVFLAVESYRLL-AKQIDQ----PLHLGITEAGGA 213
              |||||  |  :  ||:|   |:|||: :: ::||||| :|| ::    |:|||:|||
Db        294 ESAMRFSDLCNENNFNNLVFSMKASNAYVMIQSYRLLVSKQYERNMMFPIHLGVTEAGFG 353

Qy        214 RSGAVKSAIGLGLLLSEGIGDTLRVSLAADPVEEIKVGFDILKSLRIR 261
              :| :|| :|:| || 3|||||:|:||  || ||:     ::::|: |
Db        354 DNGRIKSYLGIGSLLYDGIGDTIRISLTEDPWEELTPCKKLVENLKKR 401


RESULT 15
AAW90883
ID   AAW90883 standard; Protein; 752 AA.
XX
AC   AAW90883;
XX
DT   07-JUL-2000  (first entry)
XX
DE   P. falciparum gcpE protein.
XX
KW   1-deoxy-D-xylulose-5-phosphate reductoisomerase; isoprenoid biosynthesis;
KW   deoxy-D-xylulose-phosphate; 1-deoxy-D-xylulose-5-phosphate synthase;
KW   gcpE; antimycotic; antibiotic; antiviral; antiparasitic; antimicrobial;
KW   bactericide; fungicide; herbicide; treatment; disease.
XX
OS   Plasmodium falciparum.
XX
PN   WO200017233-A2.
XX
PD   30-MAR-2000.
XX
PF   22-SEP-1999;   99WO-EP07055.
XX
PR   22-SEP-1998;   98DE-1043279.
PR   21-MAY-1999;   99DE-1023567.
XX
PA   (JOMA/) JOMAA H.
XX
PI   Jomaa H;
XX
DR   WPI; 2000-283543/24.
DR   N-PSDB; AAX82996.
XX
PT   New nucleic acid, useful therapeutically and to screen for e.g.
PT   antimicrobials and herbicides, encode proteins involved in isoprenoid
PT   biosynthesis by the deoxy-D-xylulose-phosphate route -
XX
PS   Claim 3; Page 41-44; 44pp; German.
XX
CC   This invention describes novel polypeptides and their analogs which are
CC   involved in isoprenoid biosynthesis by the deoxy-D-xylulose-phosphate
CC   route. The proteins described include 1-deoxy-D-xylulose-5-phosphate
CC   reductoisomerase, 1-deoxy-D-xylulose-5-phosphate synthase and the gcpE
CC   protein. The products of the invention have antimycotic, antibiotic,
CC   antiviral, antiparasitic, antimicrobial, bactericide, fungicide and
CC   herbicide activity. The encoding nucleic acid sequences of the invention
CC   are used for prevention or treatment of diseases in humans and animals
CC   and also to screen compounds enzyme inhibitory activity (potential
CC   antimicrobials, antiparasitic agents, antivirals, fungicides,
CC   bactericides and herbicides, for use in human or veterinary medicine or
CC   agriculture). This sequence represents the Plasmodium falciparum gcpE
CC   protein described in the method of the invention.
XX
SQ   Sequence   752 AA;

  Query Match             23.9%;  Score 446;  DB 21;  Length 752;
  Best Local Similarity   38.2%;  Pred. No. 7.1e-36;
  Matches  110;  Conservative   56;  Mismatches   84;  Indels   38;  Gaps    8;

Qy          9 RRKSTR-IYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVRVSVPTMD 67
              :|  || : :||| ||    ||:|:| :  | :||  | ||:  : :|||||::|  :
Db        117 KRLPTREVVIGNVKIGGNNKIAIQTMASCDTRNVEECVYQIRKCKDLGADIVRLTVQGVQ 176

Qy         68 AAEAF-----KLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGN-------- 114
               |:|      ||: : ||:||||||||: :|||  |:   : :|:|||| :
Db        177 EAQASYHIKEKLLSENVNIPLVADIHFNPKIALMAADV-FEKIRVNPGNYVDGRKKWIDK 235

Qy        115 ----------------EERIRMVVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALL 158
                             |: : ::  | |  |  |||| | |||   :   ||: ||  ::
Db        236 VYKTKEEFDEGKLFIKEKFVPLIEKCKR-LNRAIRIGTNHGSLSSRVLSYYGD-TPLGMV 293

Qy        159 ESAMRHVDHLDRLNFDQFKVSVKASDVFLAVESYRLL-AKQIDQ----PLHLGITEAGGA 213
              |||     |      ||      |:|||: :: ::||||| :|| ::    |:|||:||||
Db        294 ESAFEFSDLCIENNFYNLVFSMKASNAYVMIQSYRLLVSKQYERNMMFPIHLGVTEAGFG 353

Qy        214 RSGAVKSAIGLGLLLSEGIGDTLRVSLAADPVEEIKVGFDILKSLRIR 261
              :| :|| :|:| || :|||||:|:||  || ||:     ::::|: |
Db        354 DNGRIKSYLGIGSLLYDGIGDTIRISLTEDPWEELTPCKKLVENLKKR 401


Search completed: August 12, 2003, 09:56:03
Job time : 88 secs
```

OM protein - protein search, using sw model

Run on: August 12, 2003, 09:54:00 ; Search time 30 Seconds
(without alignments)
524.655 Million cell updates/sec

Title: US-09-921-992-78
Perfect score: 1866
Sequence: 1 MHNQAPIQRRKSTRIYVGNV.....RAKASQLDEARRIDVQQVEK 372

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1800	96.5	384	2	US-08-827-190-6	Sequence 6, Appli
2	1800	96.5	384	4	US-09-170-187-6	Sequence 6, Appli
3	1622.5	87.0	365	2	US-08-827-190-5	Sequence 5, Appli
4	1622.5	87.0	365	4	US-09-170-187-5	Sequence 5, Appli
5	1372.5	73.6	547	4	US-09-252-991A-22991	Sequence 22991, A
6	1195	64.0	378	4	US-09-328-352-7906	Sequence 7906, Ap
7	470.5	25.2	621	4	US-09-198-452A-389	Sequence 389, App
8	112.5	6.0	1058	4	US-09-252-991A-29105	Sequence 29105, A
9	111.5	6.0	578	3	US-09-066-046-6	Sequence 6, Appli
10	106.5	5.7	980	4	US-09-252-991A-30838	Sequence 30838, A
11	105.5	5.7	578	3	US-08-975-762-50	Sequence 50, Appl
12	105.5	5.7	578	3	US-09-295-028-50	Sequence 50, Appl
13	105.5	5.7	578	4	US-09-106-582-50	Sequence 50, Appl
14	101	5.4	465	4	US-09-252-991A-29387	Sequence 29387, A
15	101	5.4	523	4	US-09-612-964-2	Sequence 2, Appli
16	100.5	5.4	620	4	US-09-328-352-7730	Sequence 7730, Ap
17	100	5.4	343	3	US-09-039-859-2	Sequence 2, Appli
18	99.5	5.3	1374	4	US-09-252-991A-24636	Sequence 24636, A
19	97	5.2	1037	4	US-09-134-001C-4794	Sequence 4794, Ap
20	96.5	5.2	705	4	US-09-328-352-7436	Sequence 7436, Ap
21	96.5	5.2	851	4	US-09-252-991A-24773	Sequence 24773, A
22	95.5	5.1	303	4	US-09-134-001C-4855	Sequence 4855, Ap
23	95.5	5.1	740	1	US-08-309-512-10	Sequence 10, Appl
24	95.5	5.1	740	5	PCT-US92-08756A-10	Sequence 10, Appl
25	95	5.1	796	4	US-09-252-991A-17763	Sequence 17763, A
26	94.5	5.1	483	4	US-08-887-534A-51	Sequence 51, Appl
27	94.5	5.1	483	4	US-09-527-431-51	Sequence 51, Appl
28	94.5	5.1	659	4	US-09-252-991A-17904	Sequence 17904, A
29	94.5	5.1	1503	3	US-08-976-255-14	Sequence 14, Appl
30	93.5	5.0	619	4	US-09-252-991A-17411	Sequence 17411, A
31	93.5	5.0	659	4	US-09-252-991A-17731	Sequence 17731, A
32	93	5.0	459	4	US-09-491-785-2	Sequence 2, Appli
33	93	5.0	483	4	US-09-252-991A-27988	Sequence 27988, A
34	93	5.0	607	4	US-09-252-991A-20596	Sequence 20596, A
35	92.5	5.0	489	4	US-09-252-991A-27651	Sequence 27651, A
36	92.5	5.0	940	4	US-09-512-250C-2	Sequence 2, Appli
37	91	4.9	470	4	US-09-252-991A-26312	Sequence 26312, A
38	91	4.9	817	4	US-09-252-991A-31147	Sequence 31147, A
39	91	4.9	1181	4	US-09-252-991A-18480	Sequence 18480, A
40	90.5	4.8	484	4	US-09-252-991A-20787	Sequence 20787, A
41	90.5	4.8	489	4	US-09-252-991A-19810	Sequence 19810, A
42	90.5	4.8	573	4	US-09-252-991A-26428	Sequence 26428, A
43	90.5	4.8	1612	1	US-08-169-927-2	Sequence 2, Appli
44	90	4.8	358	4	US-09-784-508-4	Sequence 4, Appli
45	90	4.8	629	4	US-09-252-991A-31575	Sequence 31575, A

ALIGNMENTS

RESULT 1
US-08-827-190-6
; Sequence 6, Application US/08827190
; Patent No. 5858367
; GENERAL INFORMATION:
; APPLICANT: Rather, Philip N.
; TITLE OF INVENTION: Methods For Screening For Antimicrobials
; TITLE OF INVENTION: Utilizing aarC And Compositions Thereof
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/827,190
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Carroll, Peter G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: CASE-02443
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 384 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-827-190-6

Query Match 96.5%; Score 1800; DB 2; Length 384;
Best Local Similarity 96.9%; Pred. No. 4.9e-188;
Matches 372; Conservative 0; Mismatches 0; Indels 12; Gaps 6;

```
Qy          1 MHNQAPIQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATV--NQIKALERVGADI 58
              |||||||||||||||||||||||||||||||||||||||||||||  ||||||||||||
Db          1 MHNQAPIQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVFTNQIKALERVGADI 60

Qy         59 VRVSVPTMDAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDC--LRINPGNIGNEE 116
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||  ||||||||||||
Db         61 VRVSVPTMDAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCFTLRINPGNIGNEE 120

Qy        117 RIRMVVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMR--HVDHLDRLNFD 174
                ||||||||||||||||||||||||||||||||||||||||||||||  ||||||||||
Db        121 RIRMVVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRFTHVDHLDRLNFD 180

Qy        175 QFKVSVKASDVFLAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAI--GLGLLLSEGI 232
                ||||||||||||||||||||||||||||||||||||||||||||||  |||||||||
Db        181 QFKVSVKASDVFLAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIFTGLGLLLSEGI 240

Qy        233 GDTLRVSLAADPVEEIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVI--GTVNALEQR 290
                ||||||||||||||||||||||||||||||||||||||||||||||  |||||||||
Db        241 GDTLRVSLAADPVEEIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIFTGTVNALEQR 300

Qy        291 LEDIITPMDVSIIGCVVNGPGEALVSTLGVTGGNKKSGLYEDGVRKDRLD--NNDMIDQL 348
                ||||||||||||||||||||||||||||||||||||||||||||||  ||||||||
Db        301 LEDIITPMDVSIIGCVVNGPGEALVSTLGVTGGNKKSGLYEDGVRKDRLDFTNNDMIDQL 360

Qy        349 EARIRAKASQLDEARRIDVQQVEK 372
                ||||||||||||||||||||||||
Db        361 EARIRAKASQLDEARRIDVQQVEK 384
```

```
RESULT 2
US-09-170-187-6
; Sequence 6, Application US/09170187
; Patent No. 6383745
;  GENERAL INFORMATION:
;    APPLICANT:  Rather, Philip N.
;    TITLE OF INVENTION:  Methods For Screening For Antimicrobials
;    TITLE OF INVENTION:  Utilizing aarC And Compositions Thereof
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Medlen & Carroll, LLP
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/170,187
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/827,190
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Carroll, Peter G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  CASE-02443
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  384 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-170-187-6

  Query Match           96.5%;  Score 1800;  DB 4;  Length 384;
  Best Local Similarity   96.9%;  Pred. No. 4.9e-188;
  Matches  372;  Conservative    0;  Mismatches    0;  Indels   12;  Gaps    6;

Qy          1 MHNQAPIQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATV--NQIKALERVGADI 58
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||  ||||||||||||
Db          1 MHNQAPIQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVFTNQIKALERVGADI 60

Qy         59 VRVSVPTMDAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDC--LRINPGNIGNEE 116
                ||||||||||||||||||||||||||||||||||||||||||||||  ||||||||||||
Db         61 VRVSVPTMDAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCFTLRINPGNIGNEE 120

Qy        117 RIRMVVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMR--HVDHLDRLNFD 174
                ||||||||||||||||||||||||||||||||||||||||||||||  ||||||||||
Db        121 RIRMVVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRFTHVDHLDRLNFD 180

Qy        175 QFKVSVKASDVFLAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAI--GLGLLLSEGI 232
                ||||||||||||||||||||||||||||||||||||||||||||||  |||||||||
Db        181 QFKVSVKASDVFLAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIFTGLGLLLSEGI 240

Qy        233 GDTLRVSLAADPVEEIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVI--GTVNALEQR 290
                ||||||||||||||||||||||||||||||||||||||||||||||  |||||||||
Db        241 GDTLRVSLAADPVEEIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIFTGTVNALEQR 300

Qy        291 LEDIITPMDVSIIGCVVNGPGEALVSTLGVTGGNKKSGLYEDGVRKDRLD--NNDMIDQL 348
                ||||||||||||||||||||||||||||||||||||||||||||||  ||||||||
Db        301 LEDIITPMDVSIIGCVVNGPGEALVSTLGVTGGNKKSGLYEDGVRKDRLDFTNNDMIDQL 360

Qy        349 EARIRAKASQLDEARRIDVQQVEK 372
                ||||||||||||||||||||||||
Db        361 EARIRAKASQLDEARRIDVQQVEK 384
```

```
RESULT 3
US-08-827-190-5
; Sequence 5, Application US/08827190
; Patent No. 5858367
;  GENERAL INFORMATION:
;    APPLICANT:  Rather, Philip N.
;    TITLE OF INVENTION:  Methods For Screening For Antimicrobials
;    TITLE OF INVENTION:  Utilizing aarC And Compositions Thereof
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Medlen & Carroll, LLP
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/827,190
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Carroll, Peter G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  CASE-02443
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;     LENGTH: 365 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-827-190-5

  Query Match           87.0%;  Score 1622.5;  DB 2;  Length 365;
  Best Local Similarity   88.4%;  Pred. No. 1.2e-168;
  Matches  320;  Conservative   24;  Mismatches   17;  Indels    1;  Gaps    1;

Qy          1 MHNQAPIQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVR 60
```

```
Db          1 MHNESPIKRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVRQIQSLERVGVDIVR 60

Qy         61 VSVPTMDAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRM 120
Db         61 VSVPTMDAAEAFKLIKQRVNVPLVADIHFDYRIAMKVAEYGVDCLRINPGNIGSEERIRQ 120

Qy        121 VVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSV 180
Db        121 VVDSARHHNIPIRIGVNGGSLEKDIQEKYGEPTPEALVESAMRHVDILDRLNFDQFKVSV 180

Qy        181 KASDVFLAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSL 240
Db        181 KASDVFLAVGSYRLLAQKIDQPLHLGITEAGGARSGSVKSAIGLGMLLAEGIGDTLRISL 240

Qy        241 AADPVEEIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDV 300
Db        241 AADPVEEVKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDV 300

Qy        301 SIIGCVVNGPGEALVSTLGVTGGNKKSGLYEDGVR-KDRLDNNDMIDQLEARIRAKASQL 359
Db        301 SIIGCVVNGPGEAEVSTLGVAGAKTKSGFYEDGVRKKERFDNDNIIDQLEAKIRAKAAML 360

Qy        360 DE 361
Db        361 DE 362
```

RESULT 4
US-09-170-187-5
; Sequence 5, Application US/09170187
; Patent No. 6383745
; GENERAL INFORMATION:
; APPLICANT: Rather, Philip N.
; TITLE OF INVENTION: Methods For Screening For Antimicrobials
; TITLE OF INVENTION: Utilizing aarC And Compositions Thereof
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/170,187
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/827,190
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Carroll, Peter G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: CASE-02443
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 365 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-170-187-5

Query Match 87.0%; Score 1622.5; DB 4; Length 365;
Best Local Similarity 88.4%; Pred. No. 1.2e-168;
Matches 320; Conservative 24; Mismatches 17; Indels 1; Gaps 1;

```
Qy          1 MHNQAPIQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVR 60
Db          1 MHNESPIKRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVRQIQSLERVGVDIVR 60

Qy         61 VSVPTMDAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRM 120
Db         61 VSVPTMDAAEAFKLIKQRVNVPLVADIHFDYRIAMKVAEYGVDCLRINPGNIGSEERIRQ 120

Qy        121 VVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSV 180
Db        121 VVDSARHHNIPIRIGVNGGSLEKDIQEKYGEPTPEALVESAMRHVDILDRLNFDQFKVSV 180

Qy        181 KASDVFLAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSL 240
Db        181 KASDVFLAVGSYRLLAQKIDQPLHLGITEAGGARSGSVKSAIGLGMLLAEGIGDTLRISL 240

Qy        241 AADPVEEIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDV 300
Db        241 AADPVEEVKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDV 300

Qy        301 SIIGCVVNGPGEALVSTLGVTGGNKKSGLYEDGVR-KDRLDNNDMIDQLEARIRAKASQL 359
Db        301 SIIGCVVNGPGEAEVSTLGVAGAKTKSGFYEDGVRKKERFDNDNIIDQLEAKIRAKAAML 360

Qy        360 DE 361
Db        361 DE 362
```

RESULT 5
US-09-252-991A-22991
; Sequence 22991, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 22991
; LENGTH: 547
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-22991

Query Match 73.6%; Score 1372.5; DB 4; Length 547;
Best Local Similarity 73.8%; Pred. No. 4.9e-141;
Matches 267; Conservative 46; Mismatches 48; Indels 1; Gaps 1;

```
Qy          1 MHNQAPIQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVR 60
Db        179 IHSASPIIRRKSRKIWVGNVPVGGDAPIAVQSMTNTETCDVAATVAQIRRLEDAGADIVR 238

Qy         61 VSVPTMDAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRM 120
Db        239 VSVPDMDAAEAFGKIKQQVNVPLVADIHFDYRIALRVAELGVDCLRINPGNIGREDRVKA 298

Qy        121 VVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSV 180
Db        299 VVDAARERNIPIRIGVNAGSLEKDLQKKYGEPTPEALLESAMRHVDHLDKLDFQNFKVSV 358

Qy        181 KASDVFLAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSL 240
Db        359 KASDVFMAVAAYRLLARQIEQPLHLGITEAGGLRSGTVKSAVGLGMLLAEGIGDTIRISL 418
```

```
Qy        241 AADPVEEIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDV 300
              ||||||||||||||||||:||||||||||:|||| |||: |:| || ||||:: ||||
Db        419 AADPVEEIKVGFDILKSLHLRSRGINFIACPSCSRQNFDVVKTMNELEGRLEDLLVPMDV 478

Qy        301 SIIGCVVNGPGEALVSTLGVTGGNKKSGLYEDGVRKDRLDNNDMIDQLEARIRAKASQLD 360
              ::||||||||||  :|:|||     :| ||    :| |::::|:||  || ||::
Db        479 AVIGCVVNGPGEAKEAHVGLTGGTPNL-VYIDGKPSQKLTNDNLVDELERLIRQKAAEKA 537

Qy        361 EA 362
              ||
Db        538 EA 539


RESULT 6
US-09-328-352-7906
; Sequence 7906, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
;  APPLICANT: Gary L. Breton et al.
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
;  TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC99-03PA
;  CURRENT APPLICATION NUMBER: US/09/328,352
;  CURRENT FILING DATE: 1999-06-04
;  NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 7906
;  LENGTH: 378
;  TYPE: PRT
;  ORGANISM: Acinetobacter baumannii
US-09-328-352-7906

  Query Match             64.0%;  Score 1195;  DB 4;  Length 378;
  Best Local Similarity   63.8%;  Pred. No. 6.7e-122;
  Matches  231;  Conservative   65;  Mismatches   66;  Indels    0;  Gaps    0;

Qy          6 PIQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVRVSVPT 65
              ||:|| : :| ||:| :|  |||:|||||| | ||:||| ||:     ||||:|||||:
Db         12 PIKRRPTRKIRVGSVYVGGDAPISVQSMTNTETCDVDATVAQIERCVDAGADIMRVSVPS 71

Qy         66 MDAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRMVVDCA 125
              |:|||||  |:::|:|:|||||||||:|||| ||:|| |||||||||||:::::| ||  |
Db         72 MEAAEAFGAIRKRVSVPLVADIHFDHRIALAVADYGADCLRINPGNIGSDQKVREVVAAA 131

Qy        126 RDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSVKASDV 185
              |   | :|||||||||||||:|||||| |||||||:||:| ||||:| :||||||||:|
Db        132 RHHGISMRIGVNAGSLEKDLQKKYGEPTGQALLESALRHIDILDRLDFHEFKVSVKASNV 191

Qy        186 FLAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSLAADPV 245
              || :::|||::||| ||||:|||  |:| ||||| || || ||||||:|:||||:|
Db        192 FLTMDAYRLLSQQIDNPLHLGVTEAGIYRTGTVKSAIALGGLLMEGIGDTMRISLAAEPE 251

Qy        246 EEIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDVSIIGC 305
              :|||:|||||||| :|||||||||:||||||:||  : |||:||||| ||||||:|||
Db        252 DEIKIGFDILKSLGLRSNGINFIACPSCSRQEFNVIQVMQALEERLEDIRTPMDVSVIGC 311

Qy        306 VVNGPGEALVSTLGVTGGNKKSGLYEDGVRKDRLDNNDMIDQLEARIRAKASQLDEARRI 365
              |||||||  : :|| ||  :| :| :| :   :| | ::|::|  :| :  :|:||:
Db        312 KVNGPGEAKEADIGVVGAAPRSLVYRNGEKSHLIDTNQLVDEIETMVRQRVQELEEAKSK 371

Qy        366 DV 367
              ::
Db        372 EI 373


RESULT 7
US-09-198-452A-389
; Sequence 389, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
;  APPLICANT: Griffais, R.
;  TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
;  TITLE OF INVENTION:  thereof and uses thereof, in particular for the diagnosis, preve
;  TITLE OF INVENTION:  and treatment of infection
;  FILE REFERENCE: 9710-003-999
;  CURRENT APPLICATION NUMBER: US/09/198,452A
;  CURRENT FILING DATE:  1998-11-24
;  NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 389
;  LENGTH: 621
;  TYPE: PRT
;  ORGANISM: Chlamydia pneumoniae
US-09-198-452A-389

  Query Match             25.2%;  Score 470.5;  DB 4;  Length 621;
  Best Local Similarity   38.9%;  Pred. No. 2.5e-42;
  Matches  111;  Conservative   55;  Mismatches   84;  Indels   35;  Gaps    5;

Qy          9 RRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVRVSVPTMDA 68
              |||:  : :||: ||    |  |||| | |||:::|| || ||     |||||:|  :
Db         20 RRKTHTVRIGNLYIGSDHSIKTQSMTTTLTTDIDSTVEQIYALAEHNCDIVRVTVQGIKE 79

Qy         69 AEAFKLIKQQ-----VNVPLVADIHFDYRIALKVAEYGVDCLRINPGN---------- 111
              |:| : ||::     :|:|||||||  : |: ||::  | :||||||
Db         80 AQACEKIKERLIALGLNIPLVADIHFFPQAAMLVADF-ADKVRINPGNYIDKRNMFKGTK 138

Qy        112 ----------IGNEERIRMVVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLES 160
                       :  ||:   :|:  :     :|||| ||| : : :|||: | : :: |
Db        139 IYTEASYAQSLLRLEEKFAPLVEKCKRLGKAMRIGVNHGSLSERIMQKYGD-TIEGMVAS 197

Qy        161 AMRHVDHLDRLNFDQFKVSVKASDVFLAVESYRLLAKQID-----QPLHLGITEAGGARS 215
              |: ::   ::||:     |:|:|:  : | :|| ||| :|      |||||:||||
Db        198 AIEYIAVCEKLNYRDVVFSMKSSNPKIMVTAYRQLAKDLDARGWLYPLHLGVTEAGMGVD 257

Qy        216 GAVKSAIGLGLLLSEGIGDTLRVSLAADPVEEIKVGFDILKSLRI 260
              | :|||:|:| ||:||:|||:| ||   |  || |   :|:  :|
Db        258 GIIKSAVGIGTLLAEGLGDTIRCSLTGCPTTEIPVCDSLLRHTKI 302


RESULT 8
US-09-252-991A-29105
; Sequence 29105, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 29105
;  LENGTH: 1058
;  TYPE: PRT
;  ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-29105

  Query Match              6.0%;  Score 112.5;  DB 4;  Length 1058;
  Best Local Similarity   21.5%;  Pred. No. 0.0084;
  Matches   95;  Conservative   50;  Mismatches  140;  Indels  157;  Gaps   21;

Qy          6 PIQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVRVSVPT 65
              |:::  |   |  | | | |  :|:  |                 | | ||     | ::|||
Db        252 PVRKADSGNFYRVNAPTGSGKSVAMVMMA------------IDAAER--GHRVAIAVPT 296

Qy         66 M-DAAEAFKLIKQQ---VNVPL-VADIH--------------------FDYRIALKVAE 99
              : :     ::|||     |   | || :|                   :||   | |
Db        297 LVEVENTVRIIKQSAAAVGKELKVAPLHSASRIYERAEMQFEQGKTDHPYDYACLLDV-- 354

Qy        100 YGVDCLRINPGN------------IGNEERIRMVVDC--------------ARDKNIPIR 133
              |  | : : |||               ||  : :  |              | : :| :
Db        355 YSADLMAVEPGNEPCFNIRISACGDDGEESSKRLKHCPFLFRCGQTRMLSQAMEADIVV- 413
```

```
Qy        134 IGVN-----AGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSVKASDVFLA 188
Db        414 --INHHALLSGTTRIPLSDAEQFPGPRSLVE----------------ILLRTTPVFLV 453

Qy        189 VESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGD-------------- 234
Db        454 DEIDGLLKSAIDS----SVIE---LRLGNVGDNSPLLRLFNHVVGKSHIPGIDRSSLYRV 506

Qy        235 -------TLRVSLAAD---------PVEEIKVGFDILKSLRIRSRGINFIACPTCSRQEF 278
Db        507 NWALMYCTLSVSQLMNLQQERYFEWPIKE-TTWSDADDTFITQALGI------TRETLEY 559

Qy        279 DVIGTVNALEQRLEDIITPMDVSIIGCVVN---GPGEALVSTLGVTGGNKKSGLYEDGVR 335
Db        560 LFSNDLGQLPHHLQ--------SLSRNLANWRSNDGEHRLETLAITLGNLIADLSESGRL 611

Qy        336 KDRLDNNDMIDQLEARIRAKAS 357
Db        612 PERLKEHD-------RIRLKAS 626
```

RESULT 9
US-09-066-046-6
; Sequence 6, Application US/09066046A
; Patent No. 6204252
; GENERAL INFORMATION:
; APPLICANT: MURPHY, Cheryl
; STOREY, James
; BELTZ, Gerald A.
; COUGHLIN, Richard T.
; TITLE OF INVENTION: CHARACTERIZATION OF GRANULOCYTIC
; EHRLICHIA AND METHODS OF USE
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HALE AND DORR LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: United States
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/066,046A
; FILING DATE: 24-Apr-1998
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Superko, Colleen
; REGISTRATION NUMBER: 39,850
; REFERENCE/DOCKET NUMBER: 106.941.155
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 526-6000
; TELEFAX: (617) 526-5000
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 578 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-066-046-6

Query Match 6.0%; Score 111.5; DB 3; Length 578;
Best Local Similarity 18.6%; Pred. No. 0.0039;
Matches 88; Conservative 77; Mismatches 172; Indels 137; Gaps 20;

```
Qy          7 IQRRKSTRIYVGNVP---------IGDGA--------PIAVQSMTNTRT-TDVEATVNQ- 47
Db        112 VQEEEGTGMYLINAPEKAVVRFFKIEKSAAEEPQTVDPSVVESATGSGVDTQEEQEIDQE 171

Qy         48 IKALERVGADIVRVSVPTMDAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRI 107
Db        172 APAIEEVETEEQEV---ILEEGTLIDLEQPVAQVPVVAEAELP---GVEAAEAIVPSL-- 223

Qy        108 NPGNIGNEERIRMVV--------DCARDKNIPIR-----IGVNAGSLEKDLQEKYGEPTP 154
Db        224 ------EENKLQEVVVAPEAQQLESAPEVSAPAQPESTVLGVAEGDLKSEVSVEANADVP 277

Qy        155 QALLESAMRHVDHLDRLNFDQFKVSV---------------------------------- 180
Db        278 QKEVISGQQEQEIAEALEGTEAPVEVKEETEVLLKEDTLIDLEQPVAQVPVVAEAELPGV 337

Qy        181 -------------KASDVFLAVESYRL-LAKQIDQPLH-----LGITEAGGARSGAVKSA 221
Db        338 EAAEAIVPSLEENKLQEVVVAPEAQQLESAPEVSAPAQPESTVLGVTE------GDLKSE 391

Qy        222 IGL----GLLLSEGIGDTLRVSLAADPVEEIKVGFD--------ILKSLRIRSRGINFIA 269
Db        392 VSVEADAGMQQEAGISD--QETQATEEVEKVEVSVETKTEEPEVILEEGTLIDLEQPVAQ 449

Qy        270 CPTCSRQEFDVIGTVNAL-----EQRLEDIITPMDVSIIGCV--VNGPGEALVSTLGVTG 322
Db        450 VPVVAEAELPGVEAAEAIVPSLEENKLQEVVVAPEAQQLESAPEVSAPVQPESTVLGVTE 509

Qy        323 GNKKSGL---------YEDGVRKDRLDNNDMIDQLEARIRAKASQLDEARRIDV 367
Db        510 GDLKSEVSVEADAGMQQEAGISDQETQATEEVEKVEVSVEADAGMQQEL--VDV 561
```

RESULT 10
US-09-252-991A-30838
; Sequence 30838, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30838
; LENGTH: 980
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30838

Query Match 5.7%; Score 106.5; DB 4; Length 980;
Best Local Similarity 20.7%; Pred. No. 0.033;
Matches 88; Conservative 65; Mismatches 159; Indels 113; Gaps 20;

```
Qy         16 YVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVRVSVPTMDAAEAFKLI 75
Db        434 FVDVVRLPPGAPRSRTRIVADQVVDVGT-----EGIHRPVPPAMRTAQPDLQAARSLGV- 487

Qy         76 KQQVNVPLVADIHF------DYRI-----ALKVAEYGVDCLRINPGNIGNEERIRMVVDC 124
Db        488 --QVGIAVVGVVEFVEGRRLECRAVVDGEAQRVAQ------RVAPGGAAGVVHAELLVVV 539

Qy        125 ARDKNIPIRIGVNAGSLEKDLQEKYGEPT---PQALLES-------AMRHVDHLDRLNFD 174
Db        540 TAQAQLPI--------LATVLAEVLGEPAFVQPRVDAEGFAAADLVLLPVVAEGQYMAFG 591

Qy        175 QFKVSVKASDVFLAVESYRLLAKQIDQPLHLGI-----------------TEAGGARSG 216
Db        592 QGQVVLPGQAVATGIEA------RVETAILIGIGLVLAAPGLEGHQPGIERSEAGTA-AD 644
```

```
Qy        217 AVKSAIGLGLLLSEGIGDTLRVSLAADPVEEIKVGFDILKSLRIRSRGINFIACPTCSRQ 276
              |  |:|:|  :  |:|3| : :     |:  | |  :         :  ||    : |
Db        645 QVLLAVGVG--IETGVAENLVIVVVRLGPEQRAVEFQAI---------VEAIAAAQVAGQ 693

Qy        277 EFDVI---------GTVNALEQRLEDIITPMDVSIIGCVVNGPGEALVSTLG-VTGGNKK 326
                :         | 8    :|:|    |  |  :       || ||   :| |||| ::
Db        694 VAVAVSVWLIARSTGALGIDPRRIE-AAHPEAVPAVAGAAADPGAALQGLMGAVTGGEER 752

Qy        327 SGLY-----------EDGVRKDR-----LDNNDMIDQLEARI--RAKASQLDEARRIDV 367
                |            |:|  :      |: | :|||  ::  | :||    | | |:
Db        753 LGALAVAAPGEDLDHPADGLRAVQAGTRPADHLDALDQLHRQVLERGEAS----AGRADL 808

Qy        368 QQVEK 372
                |::
Db        809 DAVDQ 813


RESULT 11
US-08-975-762-50
; Sequence 50, Application US/08975762
; Patent No. 6207169
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven G.
;    APPLICANT:  Lodes, Michael J.
;    APPLICANT:  Houghton, Raymond
;    TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND          TREATMEN
;    NUMBER OF SEQUENCES:  73
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/975,762
;      FILING DATE:  21-MAR-1997
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.439
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-622-4900
;      TELEFAX:  206-682-6031
;   INFORMATION FOR SEQ ID NO:  50:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  578 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-975-762-50

  Query Match             5.7%;  Score 105.5;  DB 3;  Length 578;
  Best Local Similarity   19.5%;  Pred. No. 0.018;
  Matches   91;  Conservative   77;  Mismatches  176;  Indels  123;  Gaps   21;

Qy          7 IQRRKSTRIYVGNVP---------IGDGA--------PIAVQSMTNTRT-TDVEATVNQ- 47
              :|  : | :|: | |         |   |        |  |:| | :   |  |  ::|
Db        112 VQEEEGTGMYLINAPEKAVVRFFKIEKSAAEEPQTVDPSVVESATGSGVDTQEEQEIDQE 171

Qy         48 IKALERVGAD----------------------IVRVSVPTMDAAEAF--KLIKQQVNVP 82
                |:| |  :                     :    :| ::||||    | : ::
Db        172 APAIEEVETEEQEVILEEGTLIDLEQPVAQVPVVAEAELPGVEAAEAIVPSLEENKLQEV 231

Qy         83 LVADIHFDYRIALKV---AEYGVDCLRINPGNIGNEERIRMVVDCARDKNIPIRIGVNAG 139
              :||         | :|   |:     | :  |:: :|  :    | |: :       | :|
Db        232 VVAPEAQQLESAPEVSAPAQPESTVLGVAEGDLKSEVSVEANADVAQKE-------VISG 284

Qy        140 SLEKDLQEKY-GEPTPQALLE--SAMRHVDHLDRLNFDQFKVSV---------------- 180
              |::: |   |   |  : |     :  | |  |      :| |
Db        285 QQEQEIAEALEGTEAPVEVKEETEVLLKEDTLIDLEQPVAQVPVVAEAELPGVEAAEAIV 344

Qy        181 ------KASDVFLAVESYRL-LAKQIDQPLH-----LGITEAGGARSGAVKSAIGL---- 224
                  | :| :| |: :|  | ::  |       ||:||      | :|| : :
Db        345 PSLEENKLQEVVVAPEAQQLESAPEVSAPAQPESTVLGVTE------GDLKSEVSVEADA 398

Qy        225 GLLLSEGIGDTLRVSLAADPVEEIKVGFD--------ILKSLRIRSRGINFIACPTCSRQ 276
              |:    || |  : : | : ||:::|  :        ||:   :          | :
Db        399 GMQQEAGISD--QETQATEEVEKVEVSVETKTEEPEVILEEGTLIDLEQPVAQVPVVAEA 456

Qy        277 EFDVIGTVNAL-----EQRLEDIITPMDVSIIGCV--VNGPGEALVSTLGVTGGNKKSGL 329
              |   :    |:     | :|::::   :   :      |: | :   : |||| |: || :
Db        457 ELPGVEAAEAIVPSLEENKLQEVVVAPEAQQLESAPEVSAPVQPESTVLGVTEGDLKSEV 516

Qy        330 ---------YEDGVRKDRLDNNDMIDQLEARIRAKASQLDEARRIDV 367
                      | |:        : ::::|  : | |    |   :||
Db        517 SVEADAGMQQEAGISDQETQATEEVEKVEVSVEADAGMQQEL--VDV 561


RESULT 12
US-09-295-028-50
; Sequence 50, Application US/09295028
; Patent No. 6277381
; GENERAL INFORMATION:
;  APPLICANT: Reed, Steven G.
;  APPLICANT: Lodes, Michael J.
;  APPLICANT: Houghton, Raymond L.
;  APPLICANT: McNeill, Patricia D.
;  TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
;  TITLE OF INVENTION: AND TREATMENT OF EHRLICHIA INFECTION
;  FILE REFERENCE: 210121.439C4
;  CURRENT APPLICATION NUMBER: US/09/295,028
;  CURRENT FILING DATE: 1999-04-20
;  NUMBER OF SEQ ID NOS: 85
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 50
;   LENGTH: 578
;   TYPE: PRT
;   ORGANISM: Ehrlichia sp.
US-09-295-028-50

  Query Match             5.7%;  Score 105.5;  DB 3;  Length 578;
  Best Local Similarity   19.5%;  Pred. No. 0.018;
  Matches   91;  Conservative   77;  Mismatches  176;  Indels  123;  Gaps   21;

Qy          7 IQRRKSTRIYVGNVP---------IGDGA--------PIAVQSMTNTRT-TDVEATVNQ- 47
              :|  : | :|: | |         |   |        |  |:| | :   |  |  ::|
Db        112 VQEEEGTGMYLINAPEKAVVRFFKIEKSAAEEPQTVDPSVVESATGSGVDTQEEQEIDQE 171

Qy         48 IKALERVGAD----------------------IVRVSVPTMDAAEAF--KLIKQQVNVP 82
                |:| |  :                     :    :| ::||||    | : ::
Db        172 APAIEEVETEEQEVILEEGTLIDLEQPVAQVPVVAEAELPGVEAAEAIVPSLEENKLQEV 231

Qy         83 LVADIHFDYRIALKV---AEYGVDCLRINPGNIGNEERIRMVVDCARDKNIPIRIGVNAG 139
              :||         | :|   |:     | :  |:: :|  :    | |: :       | :|
Db        232 VVAPEAQQLESAPEVSAPAQPESTVLGVAEGDLKSEVSVEANADVAQKE-------VISG 284

Qy        140 SLEKDLQEKY-GEPTPQALLE--SAMRHVDHLDRLNFDQFKVSV---------------- 180
              |::: |   |   |  : |     :  | |  |      :| |
Db        285 QQEQEIAEALEGTEAPVEVKEETEVLLKEDTLIDLEQPVAQVPVVAEAELPGVEAAEAIV 344

Qy        181 ------KASDVFLAVESYRL-LAKQIDQPLH-----LGITEAGGARSGAVKSAIGL---- 224
                  | :| :| |: :|  | ::  |       ||:||      | :|| : :
Db        345 PSLEENKLQEVVVAPEAQQLESAPEVSAPAQPESTVLGVTE------GDLKSEVSVEADA 398

Qy        225 GLLLSEGIGDTLRVSLAADPVEEIKVGFD--------ILKSLRIRSRGINFIACPTCSRQ 276
              |:    || |  : : | : ||:::|  :        ||:   :          | :
Db        399 GMQQEAGISD--QETQATEEVEKVEVSVETKTEEPEVILEEGTLIDLEQPVAQVPVVAEA 456
```

```
Qy        277 EFDVIGTVNAL-----EQRLEDIITPMDVSIIGCV--VNGPGEALVSTLGVTGGNKKSGL 329
Db        457 ELPGVEAAEAIVPSLEENKLQEVVVAPEAQQLESAPEVSAPVQPESTVLGVTEGDLKSEV 516

Qy        330 ---------YEDGVRKDRLDNNDMIDQLEARIRAKASQLDEARRIDV 367
Db        517 SVEADAGMQQEAGISDQETQATEEVEKVEVSVEADAGMQQEL--VDV 561


RESULT 13
US-09-106-582-50
; Sequence 50, Application US/09106582
; Patent No. 6306402
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven G.
;    APPLICANT:  Lodes, Michael J.
;    APPLICANT:  Houghton, Raymond
;    TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND          THERAPY
;    NUMBER OF SEQUENCES:  73
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/106,582
;      FILING DATE:  29-JUN-1998
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.439C2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-622-4900
;      TELEFAX:  206-682-6031
;  INFORMATION FOR SEQ ID NO:  50:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH: 578 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-106-582-50

  Query Match             5.7%;  Score 105.5;  DB 4;  Length 578;
  Best Local Similarity   19.5%;  Pred. No. 0.018;
  Matches   91;  Conservative   77;  Mismatches  176;  Indels  123;  Gaps   21;

Qy          7 IQRRKSTRIYVGNVP---------IGDGA--------PIAVQSMTNTRT-TDVEATVNQ- 47
Db        112 VQEEEGTGMYLINAPEKAVVRFFKIEKSAAEEPQTVDPSVVESATGSGVDTQEEQEIDQE 171

Qy         48 IKALERVGAD----------------------IVRVSVPTMDAAEAF--KLIKQQVNVP 82
Db        172 APAIEEVETEEQEVILEEGTLIDLEQPVAQVPVVAEAELPGVEAAEAIVPSLEENKLQEV 231

Qy         83 LVADIHFDYRIALKV---AEYGVDCLRINPGNIGNEERIRMVVDCARDKNIPIRIGVNAG 139
Db        232 VVAPEAQQLESAPEVSAPAQPESTVLGVAEGDLKSEVSVEANADVAQKE-------VISG 284

Qy        140 SLEKDLQEKY-GEPTPQALLE--SAMRHVDHLDRLNFDQFKVSV---------------- 180
Db        285 QQEQEIAEALEGTEAPVEVKEETEVLLKEDTLIDLEQPVAQVPVVAEAELPGVEAAEAIV 344

Qy        181 ------KASDVFLAVESYRL-LAKQIDQPLH-----LGITEAGGARSGAVKSAIGL---- 224
Db        345 PSLEENKLQEVVVAPEAQQLESAPEVSAPAQPESTVLGVTE------GDLKSEVSVEADA 398

Qy        225 GLLLSEGIGDTLRVSLAADPVEEIKVGFD--------ILKSLRIRSRGINFIACPTCSRQ 276
Db        399 GMQQEAGISD--QETQATEEVEKVEVSVETKTEEPEVILEEGTLIDLEQPVAQVPVVAEA 456

Qy        277 EFDVIGTVNAL-----EQRLEDIITPMDVSIIGCV--VNGPGEALVSTLGVTGGNKKSGL 329
Db        457 ELPGVEAAEAIVPSLEENKLQEVVVAPEAQQLESAPEVSAPVQPESTVLGVTEGDLKSEV 516

Qy        330 ---------YEDGVRKDRLDNNDMIDQLEARIRAKASQLDEARRIDV 367
Db        517 SVEADAGMQQEAGISDQETQATEEVEKVEVSVEADAGMQQEL--VDV 561


RESULT 14
US-09-252-991A-29387
; Sequence 29387, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 29387
;   LENGTH: 465
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-29387

  Query Match             5.4%;  Score 101;  DB 4;  Length 465;
  Best Local Similarity   21.8%;  Pred. No. 0.038;
  Matches   81;  Conservative   43;  Mismatches  122;  Indels  126;  Gaps   15;

Qy         25 GAPIAV-QSMTNTRTTDVEATVNQIKALERVGADIVRVSVPTMDAAEAFKLIKQQVNVPL 83
Db         59 GLEIALAETLVALALDDLE----EDRADHRVGEDL------QQDAVVARGAVDQQVQFAQ 108

Qy         84 VADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRMVVDCA---------------- 125
Db        109 LLE-----RLAMP-GDAGVDLLVV--GVRGRHELHAVVAQLAHGGIDVVGGQGDVLDAFA 160

Qy        126 --------------------RDKNIPIR----IGVNAGSLEKDLQEKYGEPTPQALLESA- 161
Db        161 MVFADEFLDLRGIVGRFVDRDADLPARCGHGAGEQAGELAFDVEVANLAEVGDALVETGP 220

Qy        162 ------------MRHVDHLDRL-----NFDQFKVSVKASDVFLAVESYRLLAKQIDQPLH 204
Db        221 DVHLAALDVVGQVVDVGQADRVVVGAAAFDELEVDVIDAAVAVAVDEIQQRTADAFQAGD 280

Qy        205 LGITEAGGA--RSGAVKSAIGLGL---LLSEGIGDTLRVSLAADPVEEIKVGFDILKSLR 259
Db        281 VQLAEVGVAAHQLGTLGFGVGGGLGSVLHPEGHGAGARAVLEGE---------------- 324

Qy        260 IRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDVSIIGCVVNGPGEA------ 313
Db        325 -------FVDMPGRAAVEHDV------------DIVLLEQEDFLGTMLGGFGEAHHVEQL 365

Qy        314 --LVSTLGVTGG 323
Db        366 AQLLDALGVGGG 377


RESULT 15
US-09-612-964-2
```

; Sequence 2, Application US/09612964
; Patent No. 6403342
; GENERAL INFORMATION:
; APPLICANT: Gusyatiner Mikhail Markovich
; APPLICANT: Lunts Maria Grigorievna
; APPLICANT: Kozlov Yury Ivanovich
; APPLICANT: Ivanovskaya Lirina Valerievna
; APPLICANT: Voroshilova Elvira Borisovna
; TITLE OF INVENTION: DNA CODING FOR MUTANT ISOPROPYLMALATE SYNTHASE,
; TITLE OF INVENTION: L-LEUCINE-PRODUCING MICROORGANISM AND METHOD FOR PRODUCING
; TITLE OF INVENTION: L-LEUCINE
; FILE REFERENCE: 193845US0
; CURRENT APPLICATION NUMBER: US/09/612,964
; CURRENT FILING DATE: 2000-07-10
; PRIOR APPLICATION NUMBER: RU 99114325
; PRIOR FILING DATE: 1999-07-09
; NUMBER OF SEQ ID NOS: 12
; SEQ ID NO 2
; LENGTH: 523
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-612-964-2

Query Match 5.4%; Score 101; DB 4; Length 523;
Best Local Similarity 20.3%; Pred. No. 0.046;
Matches 83; Conservative 62; Mismatches 142; Indels 122; Gaps 18;

```
Qy          12 STRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVRVSVPTMDAA-- 69
               | :: : :  : ||    |::      |:  :    ||||:| |:: |  |
Db           2 SQQVIIFDTTLRDGEQALQASLS------VKEKLQIALALERMGVDVMEVGFPVSSPGDF 55

Qy          70 EAFKLIKQQVN----VPLVADIHFDYRIA---LKVAEYGVDCLRI------NPGNIGNE- 115
               |: : | :||      ||   :  |  :|   |||||      ||      :| :|  :
Db          56 ESVQTIARQVKNSRVCALARCVEKDIDVAAESLKVAE----AFRIHTFIATSPMHIATKL 111

Qy         116 --------ERIRMVVDCAR------------------DKNIPIRIGVNAGSLEKDLQEK 148
                     ||   :| ||                   |    :   :|||:   :: :
Db         112 RSTLDEVIERAIYMVKRARNYTDDVEFSCEDAGRTPIADLARVVEAAINAGATTINIPDT 171

Qy         149 YGEPTP---QALLESAMRHVDHLDRLNFDQFKVSVKA-SDVFLAVESYRLLAKQIDQPLH 204
                |   |     ::    | |      | |:  :||    |: ||| :  | |      :|
Db         172 VGYTMPFEFAGIISGLYERVP-----NIDKAIISVHTHDDLGLAVGN-SLAA------VH 219

Qy         205 LGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSLAADPVEEIKVGFDILKSLRIRSRG 264
                |  :  || :       .      |||:        : :  |||  |||
Db         220 AGARQVEGAMN-----------GIGERAGNCSLEEVIMAIKVRKDIL--------- 255

Qy         265 INFIACPTCSRQEFDVIGTVNALE-QRLEDIITPMDVSIIGCVVNGPGEALVSTLGVTGG 323
                        :|  :|  |  |   :::         : | |  |  : :|
Db         256 -------------NVHTAINHQEIWRTSQLVSQ--------ICNMPIPANKAIVGSGAF 293

Qy         324 NKKSGLYEDGVRKDRLDNNDM------IDQLEARIRAKASQLDEARRID 366
                  ||:::||| |:| 3   |      ::|::  : ::: :     |:|
Db         294 AHSSGIHQDGVLKNRENYEIMTPESIGLNQIQLNLTSRSGRAAVKHRMD 342
```

Search completed: August 12, 2003, 09:59:12
Job time : 32 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 12, 2003, 09:56:55 ; Search time 56 Seconds
(without alignments)
845.604 Million cell updates/sec

Title: US-09-921-992-78
Perfect score: 1866
Sequence: 1 MHNQAPIQRRKSTRIYVGNV.....RAKASQLDEARRIDVQQVEK 372

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 479057 seqs, 127295195 residues

Total number of hits satisfying chosen parameters: 479057

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	808	43.3	385	15	US-10-156-761-10098	Sequence 10098, A
2	794	42.6	359	10	US-09-881-752A-196	Sequence 196, App
3	786.5	42.1	385	15	US-10-156-761-9186	Sequence 9186, Ap
4	761	40.8	378	10	US-09-738-626-5705	Sequence 5705, Ap
5	751.5	40.3	387	10	US-09-712-363-256	Sequence 256, App
6	108	5.8	1046	15	US-10-156-761-10088	Sequence 10088, A
7	105.5	5.7	578	9	US-09-159-469-50	Sequence 50, Appl
8	105.5	5.7	578	9	US-09-798-042-50	Sequence 50, Appl
9	99	5.3	2368	9	US-09-815-242-5635	Sequence 5635, Ap
10	99	5.3	2368	9	US-09-815-242-12389	Sequence 12389, A
11	98.5	5.3	365	15	US-10-156-761-8947	Sequence 8947, Ap
12	96	5.1	504	9	US-09-815-242-10969	Sequence 10969, A
13	96	5.1	504	15	US-10-260-877-70	Sequence 70, Appl
14	94.5	5.1	290	9	US-09-815-242-5590	Sequence 5590, Ap
15	94.5	5.1	295	9	US-09-815-242-12428	Sequence 12428, A
16	93	5.0	459	9	US-09-815-242-12703	Sequence 12703, A
17	93	5.0	1004	10	US-09-738-626-5676	Sequence 5676, Ap
18	93	5.0	2478	9	US-09-815-242-5816	Sequence 5816, Ap
19	93	5.0	2478	9	US-09-815-242-12967	Sequence 12967, A
20	92.5	5.0	307	15	US-10-128-714-4018	Sequence 4018, Ap
21	92	4.9	291	9	US-09-815-242-11266	Sequence 11266, A
22	92	4.9	498	10	US-09-738-626-5600	Sequence 5600, Ap
23	92	4.9	1798	10	US-09-845-583-8	Sequence 8, Appli
24	91.5	4.9	297	9	US-09-815-242-13258	Sequence 13258, A
25	91.5	4.9	332	9	US-09-815-242-13072	Sequence 13072, A
26	91.5	4.9	756	10	US-09-738-626-6091	Sequence 6091, Ap
27	91	4.9	652	15	US-10-102-806-667	Sequence 667, App
28	91	4.9	855	15	US-10-156-761-14612	Sequence 14612, A
29	91	4.9	1798	10	US-09-938-275-9	Sequence 9, Appli
30	90.5	4.8	455	9	US-09-815-242-5420	Sequence 5420, Ap
31	90.5	4.8	455	9	US-09-815-242-12128	Sequence 12128, A
32	90.5	4.8	477	9	US-09-815-242-11826	Sequence 11826, A
33	90.5	4.8	502	15	US-10-156-761-12534	Sequence 12534, A
34	90.5	4.8	584	15	US-10-156-761-13475	Sequence 13475, A
35	90.5	4.8	842	9	US-09-815-242-11950	Sequence 11950, A
36	90	4.8	293	10	US-09-971-536-71	Sequence 71, Appl
37	90	4.8	358	9	US-09-784-508-4	Sequence 4, Appli
38	90	4.8	517	15	US-10-156-761-9172	Sequence 9172, Ap
39	89.5	4.8	722	15	US-10-128-714-8084	Sequence 8084, Ap
40	89.5	4.8	1343	9	US-09-815-242-11037	Sequence 11037, A
41	89	4.8	371	10	US-09-738-626-5880	Sequence 5880, Ap
42	89	4.8	531	9	US-09-815-242-5114	Sequence 5114, Ap
43	89	4.8	796	15	US-10-028-392-5	Sequence 5, Appli
44	89	4.8	906	9	US-09-905-983-54	Sequence 54, Appl
45	88.5	4.7	317	10	US-09-738-626-4379	Sequence 4379, Ap

ALIGNMENTS

RESULT 1
US-10-156-761-10098
; Sequence 10098, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 10098
; LENGTH: 385
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-10098

Query Match 43.3%; Score 808; DB 15; Length 385;
Best Local Similarity 44.9%; Pred. No. 1.9e-74;
Matches 164; Conservative 70; Mismatches 119; Indels 12; Gaps 3;

```
Qy          8 QRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVRVSVPTMD 67
              :|||| :| || | :|  ||::||||| |||:|: ||: ||  |   | ||||: || |
Db         17 ERRKSRQIQVGTVAVGGDAPVSVQSMTTTRTSDIGATLQQIAELTASGCQIVRVACPTQD 76

Qy         68 AAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGN-EERIRMVVDCAR 126
                |:|  :| ::  :|::|||||  :      | |   :|:|||||   ::::: :   |:
Db         77 DADALAVIARKSQIPVIADIHFQPKYVFAAIEAGCAAVRVNPGNIKQFDDKVKEIAKAAK 136
```

```
Qy        127 DKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSVKASDVF 186
              :   |||||||||||:: | ||||: ||:||:|||:          :  :|    |:||| :|
Db        137 EHGTPIRIGVNAGSLDRRLLEKYGKATPEALVESALWEASLFEEHDFRDIKISVKHNDPV 196

Qy        187 LAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSLAADPVE 246
              : | :|| || |  | ||||:|||| |    | :|||:   | |||||||||:||||:| |||
Db        197 VMVNAYRQLAAQCDYPLHLGVTEAGPAFQGTIKSAVAFGALLSEGIGDTIRVSLSAPPVE 256

Qy        247 EIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDVSIIGCV 306
              |||||  ||:|| :| ||||:   ::||:| | :.||         :     || :   |: |:::|||
Db        257 EIKVGIQILESLNLRQRGLEIVSCPSCGRAQVDVYKLAEEVTAGLEGMEVPLRVAVMGCV 316

Qy        307 VNGPGEALVSTLGVTGGNKKSGLYEDGVRKDRLDNNDMIDQLEARIRAKASQLDEARRID 366
              |||||||   :  |||   |||  |  ::   |        :   :  :::  |              ::|| :|
Db        317 VNGPGEAREADLGVASGNGKGQIFVKGEIIKTVPESKIVETL----------IEEAMKI- 365

Qy        367 VQQVE 371
              :|:|
Db        366 AEQME 370
```

RESULT 2
US-09-881-752A-196
; Sequence 196, Application US/09881752A
; Patent No. US20020115078A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Oomen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Encoding No. US20020115078A1el Helicobacter Polypeptides in the
; TITLE OF INVENTION: Genome
; FILE REFERENCE: 06132/041002
; CURRENT APPLICATION NUMBER: US/09/881,752A
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 08/833,457
; PRIOR FILING DATE: 1997-04-01
; NUMBER OF SEQ ID NOS: 370
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 196
;   LENGTH: 359
;   TYPE: PRT
;   ORGANISM: Helicobacter pylori
US-09-881-752A-196

```
  Query Match             42.6%;  Score 794;  DB 10;  Length 359;
  Best Local Similarity   45.8%;  Pred. No. 4.7e-73;
  Matches  162;  Conservative   73;  Mismatches  117;  Indels    2;  Gaps    2;

Qy          9 RRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVRVSVPTMDA 68
              | |: :|:|:| | || |||||: |||| ::| |:|:| |||   |:   |||:|||:|
Db          5 RVKTKQIFIGGVAIGGDAPISTQSMTFSKTADIESTKNQIDRLKLAGADLVRVAVSNEKD 64

Qy         69 AEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRMVVDCARDK 128
              | | | :|:    :||:|||||| |: || :|     || :|||||||||::|:|:| |||   ::|
Db         65 ALALKELKKVSPLPLIADIHFHYKFAL-IAAQSVDAIRINPGNIGSKEKIKAVVDACKEK 123

Qy        129 NIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSVKASDVFLA 188
              ||||||||||||||| |:||| |||: ::|||: :     |: |:|   ||:|:|||||
Db        124 NIPIRIGVNAGSLEKQEDQKYG-PTPKGMVESALYNAKLLEDLDFTNFKISLKASDVIRT 182

Qy        189 VESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSLAADPVEEI 248
              :|:||:|     :   | |||:||||     | ::|||:  || ||  |||||:|||:    :   ||
Db        183 IEAYRMLRPLVIYPFHLGVTEAGNLFSSSIKSAMALGGLLMEGIGDTMRVSITGELENEI 242

Qy        249 KVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDVSIIGCVVN 308
              ||    ||:          ||||:|:|||| | | :::       :|:||   | ||:|:|::|||||
Db        243 KVARAILRHSGRLKEGINWISCPTCGRIEANLVDMAIKVEKRLSHIKTPLDISVMGCVVN 302

Qy        309 GPGEALVSTLGVTGGNKKSGLYEDGVRKDRLDNNDMIDQLEARIRAKASQLDEA 362
              ||| : : :  ||:   : ::|   :|   |: :        :   | : :::
Db        303 ALGEAKHADMAIAFGNRSGLIIKEGKVIHKLAEKDLFETFVIEVENLAKEREKS 356
```

RESULT 3
US-10-156-761-9186
; Sequence 9186, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 9186
;   LENGTH: 385
;   TYPE: PRT
;   ORGANISM: Streptomyces avermitilis
US-10-156-761-9186

```
  Query Match             42.1%;  Score 786.5;  DB 15;  Length 385;
  Best Local Similarity   44.4%;  Pred. No. 3.1e-72;
  Matches  164;  Conservative   71;  Mismatches  121;  Indels   13;  Gaps    4;

Qy          6 PI-QRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVRVSVP 64
              || :|| | :| || : :|  ||::|||| |||:|: ||: ||  |   |   ||||: |
Db         15 PIAERRVSRQIQVGPLAVGGTAPVSVQSMTTTRTSDIGATLQQIAELTASGCQIVRVACP 74

Qy         65 TMDAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGN-EERIRMVVD 123
              | | |:|   :| ::   :|:||||||  :          | |   :|:|||||    ::::: :
Db         75 TQDDADALAVIARKSQIPVVADIHFQPKYVFAAIEAGCAAVRVNPGNIKQFDDKVKEIAR 134

Qy        124 CARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSVKAS 183
              |::    ||||||||||:: | ||||: ||:||:|||:        :   :|    |:||| :
Db        135 AAKEHGTPIRIGVNAGSLDRRLLEKYGKATPEALVESALWEASLFEEHDFRDIKISVKHN 194

Qy        184 DVFLAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSLAAD 243
              |  : :|| ||  |  | ||||:||||  |    | :|||:  | |||||||||:|||:|
Db        195 DPVVMVNAYRQLAAQCDYPLHLGVTEAGPAFQGTIKSAVAFGALLSEGIGDTIRVSLSAP 254

Qy        244 PVEEIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDVSII 303
              ||||||||  ||:||  :| |  :   ::||:| | :  ||          :       || :  |: |:::
Db        255 PVEEIKVGIQILESLGLRQRRLEIVSCPSCGRAQVDVYKLAEEVTAGLEGMEVPLRVAVM 314

Qy        304 GCVVNGPGEALVSTLGVTGGNKKSGLYEDGVRKDRLDNNDMIDQLEARIRAKASQLDEAR 363
              |||||||||    : |||   || |   ::   |        :   :  ::: |              ::||
Db        315 GCVVNGPGEAREADLGVASGNGKGQIFVKGEVIKTVPESKIVETL----------IEEAM 364

Qy        364 RIDVQQVEK 372
              :|   :|:|:
Db        365 KI-AEQMEQ 372
```

RESULT 4
US-09-738-626-5705
; Sequence 5705, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO

```
;  APPLICANT:  OCHIAI, KEIKO
;  APPLICANT:  YOKOI, HARUHIKO
;  APPLICANT:  TATEISHI, NAOKO
;  APPLICANT:  SENOH, AKIHIRO
;  APPLICANT:  IKEDA, MASATO
;  APPLICANT:  OZAKI, AKIO
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-125
;  CURRENT APPLICATION NUMBER: US/09/738,626
;  CURRENT FILING DATE:  2000-12-18
;  PRIOR APPLICATION NUMBER: JP 99/377484
;  PRIOR FILING DATE: 1999-12-16
;  PRIOR APPLICATION NUMBER: JP 00/159162
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: JP 00/280988
;  PRIOR FILING DATE: 2000-08-03
;  NUMBER OF SEQ ID NOS: 7059
;  SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 5705
;   LENGTH: 378
;   TYPE: PRT
;   ORGANISM: Corynebacterium glutamicum
US-09-738-626-5705

  Query Match           40.8%;  Score 761;  DB 10;  Length 378;
  Best Local Similarity   44.3%;  Pred. No. 1.3e-69;
  Matches  159;  Conservative   64;  Mismatches  124;  Indels   12;  Gaps    3;

Qy          9 RRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVRVSVPTMDA 68
              ||| : :: || | :|   ||:|||| |:| |:  |: ||  |   | |||||: |
Db          5 RRKTRQLMVGKVGVGSDHPISVQSMTTTKTHDINGTLQQIAQLTATGCDIVRVACPKTVD 64

Qy         69 AEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGN-EERIRMVVDCARD 127
              |||  :| ::  :|::|||||  :      : |   :|:|||||   : |:: |   | |
Db         65 AEALPIIAKKSPIPVIADIHFQPKYIFAAIDAGCAAVRVNPGNIKEFDGRVKEVAKAAGD 124

Qy        128 KNIPIRIGVNAGSLEKDLQEKY-GEPTPQALLESAMRHVDHLDRLNFDQFKVSVKASDVF 186
              |.|||||| |||:| : :|| |: ||:||:|||      :   |    :||| ||
Db        125 AGIPIRIGVNGGSLDKRILDKYHGKATPEALVESAMWEAGLFEEHGFGDIAISVKHSDPV 184

Qy        187 LAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSLAADPVE 246
              | ||:|| ||:| | |||||:||||    | :||::  | |||:|||||:||||:|||||
Db        185 LMVEAYRQLAEQSDYPLHLGVTEAGPKFMGTIKSSVAFGALLSQGIGDTIRVSLSADPVE 244

Qy        247 EIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDVSIIGCV 306
              |||||  ||:|| :| | :  ::||:| | : ||      : : |: :  |: |:::|||
Db        245 EIKVGDQILQSLNLRPRKLEIVSCPSCGRAQVDVYSLAEEVTEALDGMEVPLRVAVMGCV 304

Qy        307 VNGPGEALVSTLGVTGGNKKSGLYEDGVRKDRLDNNDMIDQLEARIRAKASQLDEARRI 365
              |||||||  : |||  || |  ::  |     :  : ::: |          ::|| ||
Db        305 VNGPGEARDADLGVASGNGKGQIFVKGEVIKTVPESQIVETL----------IEEAMRI 353


RESULT 5
US-09-712-363-256
; Sequence 256, Application US/09712363
; Patent No. US20020164588A1
; GENERAL INFORMATION:
;  APPLICANT: Eisenberg, David
;  APPLICANT:  Rotstein, Sergio H.
;  APPLICANT:  Marcotte, Edward M.
;  TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
;  TITLE OF INVENTION:  INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS
;  FILE REFERENCE: 07419-032001
;  CURRENT APPLICATION NUMBER: US/09/712,363
;  CURRENT FILING DATE:  2000-11-13
;  PRIOR APPLICATION NUMBER: PCT/US00/02246
;  PRIOR FILING DATE: 2000-01-28
;  PRIOR APPLICATION NUMBER: 60/179,531
;  PRIOR FILING DATE: 2000-02-01
;  PRIOR APPLICATION NUMBER: 60/117,844
;  PRIOR FILING DATE: 1999-01-29
;  PRIOR APPLICATION NUMBER: 60/118,206,
;  PRIOR FILING DATE: 1999-02-01
;  PRIOR APPLICATION NUMBER: 60/126,593
;  PRIOR FILING DATE: 1999-03-26
;  PRIOR APPLICATION NUMBER: 60/134,093
;  PRIOR FILING DATE: 1999-05-14
;  PRIOR APPLICATION NUMBER: 60/134,092
;  PRIOR FILING DATE: 1999-05-14
;  PRIOR APPLICATION NUMBER: 60/165,124
;  PRIOR FILING DATE: 1999-11-12
;  PRIOR APPLICATION NUMBER: 60/165,086
;  PRIOR FILING DATE: 1999-11-12
;  NUMBER OF SEQ ID NOS: 292
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 256
;   LENGTH: 387
;   TYPE: PRT
;   ORGANISM: Mycobacterium tuberculosis
US-09-712-363-256

  Query Match           40.3%;  Score 751.5;  DB 10;  Length 387;
  Best Local Similarity   42.7%;  Pred. No. 1.2e-68;
  Matches  156;  Conservative   64;  Mismatches  134;  Indels   11;  Gaps    2;

Qy          9 RRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVRVSVPTMDA 68
              || : :: |||| :|   |::|||| |:| || :|: ||  |   | |||||: |  :
Db         18 RRATRQLMVGNVGVGSDHPVSVQSMCTTKTHDVNSTLQQIAELTAAGCDIVRVACPRQED 77

Qy         69 AEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGN-EERIRMVVDCARD 127
              |:|   | :   :|:|||||  |      : |   :|:|||||   : |:  |   |
Db         78 ADALAEIARHSQIPVVADIHFQPRYIFAAIDAGCAAVRVNPGNIKEFDGRVGEVAKAAGA 137

Qy        128 KNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSVKASDVFL 187
              ||||||||||||:|   ||||: ||:||:|||:     :   |   |:||| :|  :
Db        138 AGIPIRIGVNAGSLDKRFMEKYGKATPEALVESALWEASLFEEHGFGDIKISVKHNDPVV 197

Qy        188 AVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSLAADPVEE 247
              | :| ||| : | |||||:|||| |  | :|||:  | ||| |||||:||||:| ||||
Db        198 MVAAYELLAARCDYPLHLGVTEAGPAFQGTIKSAVAFGALLSRGIGDTIRVSLSAPPVEE 257

Qy        248 IKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDVSIIGCVV 307
              :|||  :|:|| :| | :  ::||:| | : ||    | :   |: :  |: |:::||||
Db        258 VKVGNQVLESLNLRPRSLEIVSCPSCGRAQVDVYTLANEVTAGLDGLDVPLRVAVMGCVV 317

Qy        308 NGPGEALVSTLGVTGGNKKSGLYEDGVRKDRLDNNDMIDQLEARIRAKASQLDEARRIDV 367
              ||||||  : |||  || |  ::  |     :    ::: |          ::|| |:
Db        318 NGPGEAREADLGVASGNGKGQIFVRGEVIKTVPEAQIVETL----------IEEAMRLAA 367

Qy        368 QQVEK 372
              : |:
Db        368 EMGEQ 372


RESULT 6
US-10-156-761-10088
; Sequence 10088, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
;  APPLICANT: OMURA, SATOSHI
;  APPLICANT:  IKEDA, HARUO
;  APPLICANT:  ISHIKAWA, JUN
;  APPLICANT:  HORIKAWA, HIROSHI
;  APPLICANT:  SHIBA, TADAYOSHI
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  APPLICANT:  HATTORI, MASAHIRA
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-262
;  CURRENT APPLICATION NUMBER: US/10/156,761
;  CURRENT FILING DATE:  2002-05-29
;  PRIOR APPLICATION NUMBER: JP 2001-204089
;  PRIOR FILING DATE: 2001-05-30
;  PRIOR APPLICATION NUMBER: JP 2001-272697
;  PRIOR FILING DATE: 2001-08-02
;  NUMBER OF SEQ ID NOS: 15109
```

```
; SEQ ID NO 10088
;   LENGTH: 1046
;   TYPE: PRT
;   ORGANISM: Streptomyces avermitilis
US-10-156-761-10088

  Query Match             5.8%;  Score 108;  DB 15;  Length 1046;
  Best Local Similarity   21.8%;  Pred. No. 0.094;
  Matches   86;  Conservative   48;  Mismatches  139;  Indels  122;  Gaps   18;

Qy         11 KSTRIYVGNVPIGDGA-PIAVQSMTNTRTTDVEATVNQIKALERVGADIVRVSVPTMDAA 69
              ||| | :  |   ||  |  :::: : :  ||   |  :  ::  |||      ||
Db        614 KSTDIAILVVAANDGVMPQTIEALNHAKAADVPIVV-AVNKIDVEGAD------PT---- 662

Qy         70 EAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRMVVDCARDKN 129
                 |:  |     |||:            |||  |                | ||
Db        663 ---KVRGQLTEFGLVAE-----------EYGGD---------------TMFVD------ 686

Qy        130 IPIRIGVNAGSL----------EKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKV- 178
              |  : |:|  ||           ||:    :      :||      ||||       |
Db        687 ISAKQGLNIESLLEAVVLTADASLDLRANPEQDAQGIAIES------HLDRGRGAVATVL 740

Qy        179 ----SVKASDVFLAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGD 234
                  :::  |  :  ::|  :   :|      : ||| :    |     |||    | ||
Db        741 VQRGTLRVGDTMVVGDAYGRVRAMLDDKGE-NVEEAGPSTPVLV-----LGLTNVPGAGD 794

Qy        235 TLRVSLAADPVEEIKVGFDILKSLRIRSRGINFI------------ACPTCSRQEFDVI 281
                |     |:| :   | :    | |  ||               |      || ::|
Db        795 NFLV------VDEDRTARQIAEKRAARERNANFARRGVRFSLENLDEALKAGLVQELNLI 848

Qy        282 ------GTVNALEQRL--EDIITPMDVSIIGCVVNGPGEALVSTLGVTGGNKKSGLYEDG 333
                   |:| |||  |   |:   :|: |:     | |    | : :  |
Db        849 IKGDASGSVEALESSLLQLDVGEEVDIRILH---RGVGAVTESDINLATG---------- 895

Qy        334 VRKDRLDNNDMIDQLEARIRAKASQLDEARRIDVQ 368
                  :: ::     |   :|:|: |   :||:
Db        896 -------SDAIVIGFNVRAAGRAAQMAEREGVDVR 923


RESULT 7
US-09-159-469-50
; Sequence 50, Application US/09159469
; Patent No. US20020064535A1
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven G.
;    APPLICANT:  Lodes, Michael J.
;    APPLICANT:  Houghton, Raymond
;    TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND
;    TITLE OF INVENTION:  THERAPY OF EHRLICHIA INFECTION
;    NUMBER OF SEQUENCES:  73
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/159,469
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  09/106,582
;      FILING DATE:  29-JUN-1998
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.439C2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-622-4900
;      TELEFAX:  206-682-6031
;  INFORMATION FOR SEQ ID NO:  50:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  578 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-159-469-50

  Query Match             5.7%;  Score 105.5;  DB 9;  Length 578;
  Best Local Similarity   19.5%;  Pred. No. 0.066;
  Matches   91;  Conservative   77;  Mismatches  176;  Indels  123;  Gaps   21;

Qy          7 IQRRKSTRIYVGNVP---------IGDGA--------PIAVQSMTNTRT-TDVEATVNQ- 47
                :|  : | :|: | |          |   |        |  |:| | :   |  |  ::|
Db        112 VQEEEGTGMYLINAPEKAVVRFFKIEKSAAEEPQTVDPSVVESATGSGVDTQEEQEIDQE 171

Qy         48 IKALERVGAD----------------------IVRVSVPTMDAAEAF--KLIKQQVNVP 82
                |:| |  :                      :    :| ::||||    | : ::
Db        172 APAIEEVETEEQEVILEEGTLIDLEQPVAQVPVVAEAELPGVEAAEAIVPSLEENKLQEV 231

Qy         83 LVADIHFDYRIALKV---AEYGVDCLRINPGNIGNEERIRMVVDCARDKNIPIRIGVNAG 139
                :||        | :|   |:      | :  |:: :|  :    | |: :       | :|
Db        232 VVAPEAQQLESAPEVSAPAQPESTVLGVAEGDLKSEVSVEANADVAQKE-------VISG 284

Qy        140 SLEKDLQEKY-GEPTPQALLE--SAMRHVDHLDRLNFDQFKVSV---------------- 180
                |::: |   |   |  : |    :   | |  |     :| |
Db        285 QQEQEIAEALEGTEAPVEVKEETEVLLKEDTLIDLEQPVAQVPVVAEAELPGVEAAEAIV 344

Qy        181 ------KASDVFLAVESYRL-LAKQIDQPLH-----LGITEAGGARSGAVKSAIGL---- 224
                   |  :| :| |: :|  | ::  |       ||:||      | :|| : :
Db        345 PSLEENKLQEVVVAPEAQQLESAPEVSAPAQPESTVLGVTE------GDLKSEVSVEADA 398

Qy        225 GLLLSEGIGDTLRVSLAADPVEEIKVGFD--------ILKSLRIRSRGINFIACPTCSRQ 276
                |:    || |  : : | : ||:::|  :        ||:   :          |  :
Db        399 GMQQEAGISD--QETQATEEVEKVEVSVETKTEEPEVILEEGTLIDLEQPVAQVPVVAEA 456

Qy        277 EFDVIGTVNAL-----EQRLEDIITPMDVSIIGCV--VNGPGEALVSTLGVTGGNKKSGL 329
                |   :    |:     | :|::::   :   :     |: | :   : ||||  |: || :
Db        457 ELPGVEAAEAIVPSLEENKLQEVVVAPEAQQLESAPEVSAPVQPESTVLGVTEGDLKSEV 516

Qy        330 ---------YEDGVRKDRLDNNDMIDQLEARIRAKASQLDEARRIDV 367
                      | |:        : ::::|  : | |     |   :||
Db        517 SVEADAGMQQEAGISDQETQATEEVEKVEVSVEADAGMQQEL--VDV 561


RESULT 8
US-09-798-042-50
; Sequence 50, Application US/09798042
; Patent No. US20020068343A1
; GENERAL INFORMATION:
;  APPLICANT:  Reed, Steven G.
;  APPLICANT:  Lodes, Michael J.
;  APPLICANT:  Houghton, Raymond L.
;  APPLICANT:  McNeill, Patricia D.
;  TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR THE DIAGNOSIS
;  TITLE OF INVENTION:  AND TREATMENT OF EHRLICHIA INFECTION
;  FILE REFERENCE:  210121.439C7
;  CURRENT APPLICATION NUMBER: US/09/798,042
;  CURRENT FILING DATE:  2001-03-02
;  NUMBER OF SEQ ID NOS:  108
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 50
;   LENGTH: 578
;   TYPE: PRT
;   ORGANISM: Ehrlichia sp.
US-09-798-042-50

  Query Match             5.7%;  Score 105.5;  DB 9;  Length 578;
  Best Local Similarity   19.5%;  Pred. No. 0.066;
```

```
  Matches   91;  Conservative   77;  Mismatches  176;  Indels  123;  Gaps   21;

Qy          7 IQRRKSTRIYVGNVP---------IGDGA--------PIAVQSMTNTRT-TDVEATVNQ- 47
              :|  : | :|: | |          |   |        |  |:| | :   |  |  ::|
Db        112 VQEEEGTGMYLINAPEKAVVRFFKIEKSAAEEPQTVDPSVVESATGSGVDTQEEQEIDQE 171

Qy         48 IKALERVGAD----------------------IVRVSVPTMDAAEAF--KLIKQQVNVP 82
                |:| |  :                    :    :| ::|||||    | : ::
Db        172 APAIEEVETEEQEVILEEGTLIDLEQPVAQVPVVAEAELPGVEAAEAIVPSLEENKLQEV 231

Qy         83 LVADIHFDYRIALKV---AEYGVDCLRINPGNIGNEERIRMVVDCARDKNIPIRIGVNAG 139
              :||         | :|   |:     | :  |:: :|  :    | |: :        | :|
Db        232 VVAPEAQQLESAPEVSAPAQPESTVLGVAEGDLKSEVSVEANADVAQKE-------VISG 284

Qy        140 SLEKDLQEKY-GEPTPQALLE--SAMRHVDHLDRLNFDQFKVSV---------------- 180
              |::: |   |   | : |    :   | |  |      :| |
Db        285 QQEQEIAEALEGTEAPVEVKEETEVLLKEDTLIDLEQPVAQVPVVAEAELPGVEAAEAIV 344

Qy        181 ------KASDVFLAVESYRL-LAKQIDQPLH-----LGITEAGGARSGAVKSAIGL---- 224
                  |  :| :| |: :|  | ::  |        ||:||      | :|| : :
Db        345 PSLEENKLQEVVVAPEAQQLESAPEVSAPAQPESTVLGVTE------GDLKSEVSVEADA 398

Qy        225 GLLLSEGIGDTLRVSLAADPVEEIKVGFD--------ILKSLRIRSRGINFIACPTCSRQ 276
              |:    || |  : : | : ||:::|  :         ||:   :           |  :
Db        399 GMQQEAGISD--QETQATEEVEKVEVSVETKTEEPEVILEEGTLIDLEQPVAQVPVVAEA 456

Qy        277 EFDVIGTVNAL-----EQRLEDIITPMDVSIIGCV--VNGPGEALVSTLGVTGGNKKSGL 329
              |   :    |:     | :|:::   :   :     |: | :   : |||| |: || :
Db        457 ELPGVEAAEAIVPSLEENKLQEVVVAPEAQQLESAPEVSAPVQPESTVLGVTEGDLKSEV 516

Qy        330 ---------YEDGVRKDRLDNNDMIDQLEARIRAKASQLDEARRIDV 367
                      | |:        : ::::|  : | |    |   :||
Db        517 SVEADAGMQQEAGISDQETQATEEVEKVEVSVEADAGMQQEL--VDV 561


RESULT 9
US-09-815-242-5635
; Sequence 5635, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5635
;   LENGTH: 2368
;   TYPE: PRT
;   ORGANISM: Staphylococcus aureus

US-09-815-242-5635

  Query Match             5.3%;  Score 99;  DB 9;  Length 2368;
  Best Local Similarity   20.8%;  Pred. No. 2.9;
  Matches   89;  Conservative   58;  Mismatches  155;  Indels  126;  Gaps   19;

Qy          3 NQAPIQ-RRKSTRIYVGNVPIGDG---------------APIAVQSMTNTRTTDVE--- 42
              |||     : ||  || |  : :|                |   :::  | : |: :
Db       1532 NQAKTNIDQSSTDEYVDNA-VKEGKAKINAVKTFSEYKKDALAKIEAAYNAKVTEADNSN 1590

Qy         43 -ATVNQIKALERVGADIVRVSVPTMDAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYG 101
               :| ::|   ::  |:: : :   :: | :    |  : | :  |:
Db       1591 ASTSSEIAEAKQKLAELKQTADQNVNQATS----KDDIEVQIHNDLD------------- 1633

Qy        102 VDCLRINPGNIGNEERIRMVVDC---ARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALL 158
                 ||   |   ::      |    | |   |    ||   ||    |  :   | |
Db       1634 ----NINDYTIPTGKKESATTDLYAYADQKKNNISADTNATQDEKQQAIKQVDQNVQTAL 1689

Qy        159 ESAMRHVDH-------------LDRLNFD---------------QFKVSVKASDVFLAV 189
              ||    ||:             :| :  |                 : | |:  ||   |
Db       1690 ESINNGVDNGDVDDALTQGKAAIDAIQVDATVKPKANQVIDAKAEETKESIDQSDQLTAE 1749

Qy        190 ESYRLLA--KQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIG-DTLRVSLAADPVE 246
              |    ||  |||        |||:|   :  |: |   ||   : |  |:     |   :|
Db       1750 EKTEALAMIKQITDQAKQGITDA--TTTAEVEKAKAQGLEAFDNIQIDSTEKQKA---IE 1804

Qy        247 EIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDVSIIGCV 306
              |::   | :::      |:|  |  |   :|        |    ||||::     |
Db       1805 ELETALDQIEA------GVNVDADATTEEKE--------AFTNALEDILSKATEDISDQT 1850

Qy        307 VNGPGEALVSTLGVTGGNKKSGLYEDGVRKDRLDNNDMIDQLEA-RIR--AKASQLDEAR 363
               |    | ::|:                        |  ::||:| ||    | : |:  |
Db       1851 TN----AEIATV--------------------KNSALEQLKAQRINPVVKKNALEAIR 1884

Qy        364 RIDVQQVE 371
                :  :|:|
Db       1885 EVVNKQIE 1892


RESULT 10
US-09-815-242-12389
; Sequence 12389, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
```

```
; SEQ ID NO 12389
;   LENGTH: 2368
;   TYPE: PRT
;   ORGANISM: Staphylococcus aureus
US-09-815-242-12389

  Query Match              5.3%;  Score 99;  DB 9;  Length 2368;
  Best Local Similarity   20.8%;  Pred. No. 2.9;
  Matches   89;  Conservative   58;  Mismatches  155;  Indels  126;  Gaps   19;

Qy          3 NQAPIQ-RRKSTRIYVGNVPIGDG---------------APIAVQSMTNTRTTDVE--- 42
                |||      : ||  ||||   : :|                |   :::  | : |: :
Db       1532 NQAKTNIDQSSTDEYVDNA-VKEGKAKINAVKTFSEYKKDALAKIEAAYNAKVTEADNSN 1590

Qy         43 -ATVNQIKALERVGADIVRVSVPTMDAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYG 101
               :| ::|   ::  |:8: :    :: | :     |  : | :  |:
Db       1591 ASTSSEIAEAKQKLAELKQTADQNVNQATS----KDDIEVQIHNDLD------------- 1633

Qy        102 VDCLRINPGNIGNEERIRMVVDC---ARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALL 158
                     ||   |   ::|    |     |   |   |     ||    ||     |   :    |  |
Db       1634 ----NINDYTIPTGKKESATTDLYAYADQKKNNISADTNATQDEKQQAIKQVDQNVQTAL 1689

Qy        159 ESAMRHVDH--------------LDRLNFD---------------QFKVSVKASDVFLAV 189
              ||     ||:             :| :   |                     : | |:  ||   |
Db       1690 ESINNGVDNGDVDDALTQGKAAIDAIQVDATVKPKANQVIDAKAEETKESIDQSDQLTAE 1749

Qy        190 ESYRLLA--KQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIG-DTLRVSLAADPVE 246
              |     ||  |||       |||:|     :   |: |   ||     :   |   |:       |    :|
Db       1750 EKTEALAMIKQITDQAKQGITDA--TTTAEVEKAKAQGLEAFDNIQIDSTEKQKA---IE 1804

Qy        247 EIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDVSIIGCV 306
              |::     | :::      |:|   |   |    :|          |      ||||::     |
Db       1805 ELETALDQIEA------GVNVDADATTEEKE--------AFTNALEDILSKATEDISDQT 1850

Qy        307 VNGPGEALVSTLGVTGGNKKSGLYEDGVRKDRLDNNDMIDQLEA-RIR--AKASQLDEAR 363
                |      |  ::|:                              |    ::||:| ||     |  : |:    |
Db       1851 TN----AEIATV------------------------KNSALEQLKAQRINPVVKKNALEAIR 1884

Qy        364 RIDVQQVE 371
                :   :|:|
Db       1885 EVVNKQIE 1892
```

RESULT 11
US-10-156-761-8947
; Sequence 8947, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 8947
;   LENGTH: 365
;   TYPE: PRT
;   ORGANISM: Streptomyces avermitilis
US-10-156-761-8947

```
  Query Match              5.3%;  Score 98.5;  DB 15;  Length 365;
  Best Local Similarity   20.9%;  Pred. No. 0.17;
  Matches   77;  Conservative   54;  Mismatches  138;  Indels   99;  Gaps   19;

Qy         18 GNVPIGDGAPIAVQSMTNTRTTDVEA----------TVNQIKALERVGAD--------- 57
              |: |:||||  : :  :             | |          |:::   | ||||
Db         28 GSAPVGDGAVLHIGTDEGAEGVSVFARPGAYSTLRDLVDRVFRAELVGADPFQREWLWHR 87

Qy         58 ------IVRVSVPTMDAAE-------------AFKLIKQ-QVNVPLVADIHFDYRIA-- 94
                    |  : :||:    :                    ::|:    :   :|   |         :|
Db         88 VWELDRIHELPLPTLGLIDIALWDLAGRYHGEPVWRLLGGFREAIPAYASTSTFSSVAEF 147

Qy         95 LKVAEYGVDCLRINPGNIGNEERIRMVVDCARDKNIPIRIGVNAGSLEKDLQEKY----G 150
               | ||:     || :      |        ::       |   ||   : :  :   : |       |:   |    |
Db        148 LDVAD---QCLALGYRGI----KLHAWGDARRDAELCLALRDHVG---PDVPLMYDGSAG 197

Qy        151 EPTPQAL-LESAMRHVDHLDRLNFDQFKVSVKASDVFLAVESYRLLAKQIDQPLHLGITE 209
                  | |:  |    |:       |:|           ::    ::         :: :|:  ||:  :| ||  :   |
Db        198 FDLPDAIRLGRALSEADYL------WYEEPIRE----FSISAYQRLAEAVDVPLLVAETS 247

Qy        210 AGG-------ARSGAVKSAIGLGLLLSEGIGDTLRVSLAADPVEEIKVGFDILKSLRIRS 262
               |             |:||      :    |    |    ||      :| :   ||                    ||
Db        248 DGAHMNAADFIRAGAATFGVRAGTTLRGGITGAMRTAHLADAFR-----------LRAEV 296

Qy        263 RGI---NFIACPTCSRQEF--DVIGTVNALEQRLEDIITPMDVSIIGCVVNGPGEALVST 317
               |        |      |       |       :      :: :||  : :|     |           |    ::            ||| ||
Db        297 HGSDIPNHHLCMAISNTTYYESLVTSVNVVRERHVD-----DQGLVHAPA-GPGIALPLD 350

Qy        318 LGVTGGNK 325
                |     ||:
Db        351 FGY--GNE 356
```

RESULT 12
US-09-815-242-10969
; Sequence 10969, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10969
;   LENGTH: 504
;   TYPE: PRT
;   ORGANISM: Haemophilus influenzae
US-09-815-242-10969

```
  Query Match              5.1%;  Score 96;  DB 9;  Length 504;
  Best Local Similarity   22.0%;  Pred. No. 0.51;
  Matches   85;  Conservative   59;  Mismatches  134;  Indels  108;  Gaps   21;
```

```
Qy          36 TRTTDVEATVNQIKALER---------VGADIVRVSVPTMDAAE---AFKLIKQQVNVPL 83
               ||| |  | |      | |          | | : :    :|  |      |: :| :
Db          24 TRTRD--ALVADFPGLTRDRKYGHAHIAGYDFIVIDTGGIDGTEEGVEEKMAEQSLLAID 81

Qy          84 VADIHF---DYRIALKVAEYGV-DCLRINPGNIGNEERIRMVV----------DCARDK 128
                |||     | |  |  |: |: : ||         : :| :||           ||
Db          82 EADIVLFLVDARAGLTAADIGIANYLR------QRQNKITVVVANKTDGIDADSHCAEFY 135

Qy         129 NIPI----RIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSVKASD 184
               : :    :|  : |     | |:   |          | : |  || : ::     :  |
Db         136 QLGLGEIEQIAASQGRGVTQLMEQVLAP-----FAEKMENADENDRTSEEEQDEWEQEFD 190

Qy         185 VFLAVESYRLLAKQIDQPL------HLGITEAGGARSGAVKSAIGLGLLLSEGIGD---- 234
               | : |   |:   :|: |      :: |   |       :  :|   | :  :|:
Db         191 -FDSEEDTALIDDALDEELEEEQDKNIKIAIVG-------RPNVGKSTLTNRILGEDRVV 242

Qy         235 ------TLRVSLAADPVEEIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALE 288
                   | | |:   |:|     : :: :  :| ||   :|      ::| || |: |::
Db         243 VFDMPGTTRDSIYI-PMERDGQQYTLIDTAGVRKRGKVHLAV-----EKFSVIKTLQAIQ 296

Qy         289 QR--------LEDIITPMDVSIIGCVVNGPGEALVSTLGVTGGNKKSGLYEDGVRKDRLD 340
                        : |:  |:|:::| ::|  | :||  :     ||  || :|   |||:
Db         297 DANVVLLTIDARENISDQDLSLLGFILNA-GRSLVIVV-----NKWDGLDQD--VKDRV- 347

Qy         341 NNDMIDQLEARIRAKASQLDEARRID 366
                           |:||  ||:|
Db         348 --------------KSELD--RRLD 356
```

RESULT 13
US-10-260-877-70
; Sequence 70, Application US/10260877
; Publication No. US20030021813A1
; GENERAL INFORMATION:
; APPLICANT: Abbott Laboratories
; APPLICANT: Chovan, Linda E.
; APPLICANT: Hessler, Paul E.
; APPLICANT: Reich, Karl A.
; TITLE OF INVENTION: ESSENTIAL BACTERIA GENES AND GENOME
; TITLE OF INVENTION: SCANNING IN HAEMOPHILUS INFLUENZAE FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: 'ESSENTIAL GENES'
; FILE REFERENCE: 6565.US.P1
; CURRENT APPLICATION NUMBER: US/10/260,877
; CURRENT FILING DATE: 2002-09-30
; PRIOR APPLICATION NUMBER: US/09/649,145
; PRIOR FILING DATE: 2000-08-25
; NUMBER OF SEQ ID NOS: 137
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 70
; LENGTH: 504
; TYPE: PRT
; ORGANISM: H. influenzae
US-10-260-877-70

Query Match 5.1%; Score 96; DB 15; Length 504;
Best Local Similarity 22.0%; Pred. No. 0.51;
Matches 85; Conservative 59; Mismatches 134; Indels 108; Gaps 21;

```
Qy          36 TRTTDVEATVNQIKALER---------VGADIVRVSVPTMDAAE---AFKLIKQQVNVPL 83
               ||| |  | |      | |          | | : :    :|  |      |: :| :
Db          24 TRTRD--ALVADFPGLTRDRKYGHAHIAGYDFIVIDTGGIDGTEEGVEEKMAEQSLLAID 81

Qy          84 VADIHF---DYRIALKVAEYGV-DCLRINPGNIGNEERIRMVV----------DCARDK 128
                |||     | |  |  |: |: : ||         : :| :||           ||
Db          82 EADIVLFLVDARAGLTAADIGIANYLR------QRQNKITVVVANKTDGIDADSHCAEFY 135

Qy         129 NIPI----RIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSVKASD 184
               : :    :|  : |     | |:   |          | : |  || : ::     :  |
Db         136 QLGLGEIEQIAASQGRGVTQLMEQVLAP-----FAEKMENADENDRTSEEEQDEWEQEFD 190

Qy         185 VFLAVESYRLLAKQIDQPL------HLGITEAGGARSGAVKSAIGLGLLLSEGIGD---- 234
               | : |   |:   :|: |      :: |   |       :  :|   | :  :|:
Db         191 -FDSEEDTALIDDALDEELEEEQDKNIKIAIVG-------RPNVGKSTLTNRILGEDRVV 242

Qy         235 ------TLRVSLAADPVEEIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALE 288
                   | | |:   |:|     : :: :  :| ||   :|      ::| || |: |::
Db         243 VFDMPGTTRDSIYI-PMERDGQQYTLIDTAGVRKRGKVHLAV-----EKFSVIKTLQAIQ 296

Qy         289 QR--------LEDIITPMDVSIIGCVVNGPGEALVSTLGVTGGNKKSGLYEDGVRKDRLD 340
                        : |:  |:|:::| ::|  | :||  :     ||  || :|   |||:
Db         297 DANVVLLTIDARENISDQDLSLLGFILNA-GRSLVIVV-----NKWDGLDQD--VKDRV- 347

Qy         341 NNDMIDQLEARIRAKASQLDEARRID 366
                           |:||  ||:|
Db         348 --------------KSELD--RRLD 356
```

RESULT 14
US-09-815-242-5590
; Sequence 5590, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5590
; LENGTH: 290
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5590

Query Match 5.1%; Score 94.5; DB 9; Length 290;
Best Local Similarity 21.6%; Pred. No. 0.3;
Matches 63; Conservative 35; Mismatches 106; Indels 87; Gaps 12;

```
Qy          27 PIAVQSMTNTRTTDVEAT-----VNQIKALERVGADIVRVSVPTMDAAEAFKLIKQQVNV 81
               |  |: : |  :  | |      : : : || :| | :  |     | | : | | |  |
Db          60 PKIVEEVMNAVSIPVMAKARIGHITEARVLEAMGVDYIDESEVLTPADEEYHLRKDQFTV 119

Qy          82 PLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRMVVDCARDKNIPIRIGVNAGSL 141
               | |                          |:|   |               |||  | |
Db         120 PFVCGCR---------------------NLGEAAR--------------RIGEGAAML 142

Qy         142 EKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSVKASDVFLAVESYRLLAKQIDQ 201
                     |||    ::|:   | |: ::| :  :::|   |  :         || |
Db         143 RTK-----GEPGTGNIVEA----VRHMRQVNSEVSRLTVMNDDEIMT------FAKDIGA 187

Qy         202 PLHLGITEAGGARSGAVKSAIGLGLLLSEGIGD----TLRVSLAADPVEEIKVGFDILKS 257
               |  :        |  |  | |       |:       | : | || |    ||  | ||
```

```
Db        188 PYEILKQIKDNGRLPVVNFAAG-------GVATPQDAALMMELGADGV---FVGSGIFKS 237

Qy        258 LRIRSRGINFIACPTCSRQEFDVIG-------------TVN--ALEQRLED 293
                  :    |   |  |:::::||              :|  :||:|:::
Db        238 -EDPEKFAKAIVQATTHYQDYELIGRLASDLGTAMKGLDINQLSLEERMQE 287


RESULT 15
US-09-815-242-12428
; Sequence 12428, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12428
;   LENGTH: 295
;   TYPE: PRT
;   ORGANISM: Staphylococcus aureus
US-09-815-242-12428

  Query Match             5.1%;  Score 94.5;  DB 9;  Length 295;
  Best Local Similarity   21.6%;  Pred. No. 0.31;
  Matches   63;  Conservative   35;  Mismatches  106;  Indels   87;  Gaps   12;

Qy         27 PIAVQSMTNTRTTDVEAT-----VNQIKALERVGADIVRVSVPTMDAAEAFKLIKQQVNV 81
              |  |: : |  :  | |      : : : |+ :| | :  |     | | : | | |  |
Db         65 PKIVEEVMNAVSIPVMAKARIGHITEARVLEAMGVDYIDESEVLTPADEEYHLRKDQFTV 124

Qy         82 PLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRMVVDCARDKNIPIRIGVNAGSL 141
              | |                          |:|   |               |||  |  |
Db        125 PFVCGCR---------------------NLGEAAR---------------RIGEGAAML 147

Qy        142 EKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSVKASDVFLAVESYRLLAKQIDQ 201
                  |||     :|:    | |: ::| :  :::|   |  :        || |
Db        148 RTK-----GEPGTGNIVEA----VRHMRQVNSEVSRLTVMNDDEIMT------FAKDIGA 192

Qy        202 PLHLGITEAGGARSGAVKSAIGLGLLLSEGIGD----TLRVSLAADPVEEIKVGFDILKS 257
              |  :      |   |  | |     |:        | : | || |    ||  | ||
Db        193 PYEILKQIKDNGRLPVVNFAAG-------GVATPQDAALMMELGADGV---FVGSGIFKS 242

Qy        258 LRIRSRGINFIACPTCSRQEFDVIG-------------TVN--ALEQRLED 293
                  :    |   |  |:::::||              :|  :||:|:::
Db        243 -EDPEKFAKAIVQATTHYQDYELIGRLASDLGTAMKGLDINQLSLEERMQE 292


Search completed: August 12, 2003, 10:07:05
```

Job time : 58 secs

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 14, 2003, 21:47:30 ; Search time 4626 Seconds
(without alignments)
3289.749 Million cell updates/sec

Title: US-09-921-992-78
Perfect score: 1866
Sequence: 1 MHNQAPIQRRKSTRIYVGNV.....RAKASQLDEARRIDVQQVEK 372

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09921992/runat_12082003_094911_17364/app_query.fasta_1.519
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09921992_@CGN_1_1_3508_@runat_12082003_094911_17364 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		1866	100.0	1119	1	AY033515	AY033515 Escherich
2		1866	100.0	1119	6	AX036302	AX036302 Sequence
3		1866	100.0	1119	6	AX038912	AX038912 Sequence
4		1866	100.0	1119	6	AX050487	AX050487 Sequence
5		1866	100.0	1119	6	AX393943	AX393943 Sequence
6		1866	100.0	1697	1	ECGCPE	X64451 E.coli gcpE
7	c	1866	100.0	11521	1	AE005481	AE005481 Escherich
8	c	1866	100.0	13176	1	AE000338	AE000338 Escherich
9	c	1866	100.0	296827	1	AP002561	AP002561 Escherich
10	c	1866	100.0	300099	1	AE016764	AE016764 Escherich
11	c	1855	99.4	10225	1	AE015271	AE015271 Shigella
12	c	1855	99.4	290380	1	AE016987	AE016987 Shigella
13	c	1840	98.6	23647	1	AE008814	AE008814 Salmonell
14	c	1831	98.1	145050	1	AL627275	AL627275 Salmonell
15		1831	98.1	300431	1	AE016835	AE016835 Salmonell
16		1691	90.6	1010	6	AR027993	AR027993 Sequence
17		1691	90.6	1010	6	AR208328	AR208328 Sequence
18		1691	90.6	1010	6	BD084908	BD084908 Methods f
19		1645.5	88.2	301235	1	AP005075	AP005075 Vibrio pa
20	c	1631.5	87.4	301442	1	AE016798	AE016798 Vibrio vu
21		1624.5	87.1	1345	1	PSU67933	U67933 Providencia
22		1624.5	87.1	1345	6	AR027991	AR027991 Sequence
23		1624.5	87.1	1345	6	AR208326	AR208326 Sequence
24		1624.5	87.1	1345	6	BD084906	BD084906 Methods f
25		1622.5	87.0	11161	1	AE013738	AE013738 Yersinia
26	c	1622.5	87.0	204050	1	AJ414154	AJ414154 Yersinia
27		1621.5	86.9	11824	1	AE004161	AE004161 Vibrio ch
28		1603.5	85.9	10507	1	U32721	U32721 Haemophilus
29		1603.5	85.9	110000	6	AR274513_03	Continuation (4 of
30		1599.5	85.7	11268	1	AE006237	AE006237 Pasteurel
31		1593.5	85.4	1079	6	AR027992	AR027992 Sequence
32		1593.5	85.4	1079	6	AR208327	AR208327 Sequence
33		1593.5	85.4	1079	6	BD084907	BD084907 Methods f
34	c	1584.5	84.9	10607	1	AE015769	AE015769 Shewanell
35	c	1372.5	73.6	17148	1	AE004798	AE004798 Pseudomon
36	c	1353	72.5	886	6	AR027995	AR027995 Sequence
37	c	1353	72.5	886	6	AR208330	AR208330 Sequence
38	c	1353	72.5	886	6	BD084910	BD084910 Methods f
39		1341.5	71.9	300861	1	AE016777	AE016777 Pseudomon
40		1335.5	71.6	311249	1	AE016860	AE016860 Pseudomon
41		1285.5	68.9	9608	1	AE014104	AE014104 Buchnera
42		1282	68.7	110000	6	BD061520_3	Continuation (4 of
43		1282	68.7	347550	1	AP001118	AP001118 Buchnera
44		1226.5	65.7	349174	1	AB063522	AB063522 Wiggleswo
45	c	1186	63.6	4775	1	D90881	D90881 E.coli geno

ALIGNMENTS

RESULT 1

```
AY033515
LOCUS       AY033515               1119 bp    DNA     linear   BCT 26-DEC-2001
DEFINITION  Escherichia coli 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
            synthase (ispG) gene, complete cds.
ACCESSION   AY033515
VERSION     AY033515.1  GI:17978527
KEYWORDS    .
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 1119)
  AUTHORS   Hecht,S., Eisenreich,W., Adam,P., Amslinger,S., Kis,K., Bacher,A.,
            Arigoni,D. and Rohdich,F.
  TITLE     Studies on the nonmevalonate pathway to terpenes: The role of the
            GcpE (IspG) protein
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 98 (26), 14837-14842 (2001)
   PUBMED   11752431
REFERENCE   2  (bases 1 to 1119)
  AUTHORS   Rohdich,F., Hecht,S., Adam,P., Bacher,A. and Eisenreich,W.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-APR-2001) Institut fuer Organische Chemie und
            Biochemie, Lehrstuhl III, Technische Universitaet Muenchen,
            Lichtenberg Strasse 4, Garching D-85747, Germany
FEATURES             Location/Qualifiers
     source          1. .1119
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:562"
     gene            1. .1119
                     /gene="ispG"
     CDS             1. .1119
                     /gene="ispG"
                     /function="converts 2C-methyl-D-erythritol
                     2,4-cyclodiphosphate into 1-hydroxy-2-methyl-2-(E)-butenyl
                     4-diphosphate, last but one step of the
                     mevalonate-independent pathway for isoprenoid
                     biosynthesis"
                     /note="IspG"
                     /codon_start=1
                     /transl_table=11
                     /product="1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
                     synthase"
                     /protein_id="AAK53460.1"
                     /db_xref="GI:17978528"
                     /translation="MHNQAPIQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEAT
                     VNQIKALERVGADIVRVSVPTMDAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGV
                     DCLRINPGNIGNEERIRMVVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLES
                     AMRHVDHLDRLNFDQFKVSVKASDVFLAVESYRLLAKQIDQPLHLGITEAGGARSGAV
                     KSAIGLGLLLSEGIGDTLRVSLAADPVEEIKVGFDILKSLRIRSRGINFIACPTCSRQ
                     EFDVIGTVNALEQRLEDIITPMDVSIIGCVVNGPGEALVSTLGVTGGNKKSGLYEDGV
                     RKDRLDNNDMIDQLEARIRAKASQLDEARRIDVQQVEK"
BASE COUNT      272 a    280 c    315 g    252 t
ORIGIN

Alignment Scores:
Pred. No.:                9.1e-136      Length:          1119
Score:                    1866.00       Matches:         372
Percent Similarity:       100.00%       Conservative:    0
Best Local Similarity:    100.00%       Mismatches:      0
Query Match:              100.00%       Indels:          0
DB:                       1             Gaps:            0

US-09-921-992-78 (1-372) x AY033515 (1-1119)

Qy          1 MetHisAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCATAACCAGGCTCCAATTCAACGTAGAAAATCAACACGTATTTACGTTGGGAATGTG 60

Qy         21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrThrAsp 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCGATTGGCGATGGTGCTCCCATCGCCGTACAGTCCATGACCAATACGCGTACGACAGAT 120

Qy         41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GTCGAAGCAACGGTCAATCAAATCAAGGCGCTGGAACGCGTTGGCGCTGATATCGTCCGT 180

Qy         61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsn 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GTCTCCGTACCGACGATGGACGCGGCAGAAGCGTTCAAACTCATCAAACAGCAGGTTAAC 240

Qy         81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTGCCGCTGGTGGCTGACATCCACTTCGACTATCGCATTGCGCTGAAAGTAGCGGAATAC 300

Qy        101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMet 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGCGTCGATTGTCTGCGTATTAACCCTGGCAATATCGGTAATGAAGAGCGTATTCGCATG 360

Qy        121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GTGGTTGACTGTGCGCGCGATAAAAACATTCCGATCCGTATTGGCGTTAACGCCGGATCG 420

Qy        141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSer 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTGGAAAAAGATCTGCAAGAAAAGTATGGCGAACCGACGCCGCAGGCGTTGCTGGAATCC 480

Qy        161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GCCATGCGCCATGTTGATCATCTCGATCGCCTGAACTTCGATCAGTTCAAAGTCAGCGTG 540

Qy        181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AAAGCGTCTGACGTCTTCCTCGCTGTTGAGTCTTATCGTTTGCTGGCAAAACAAATCGAT 600

Qy        201 GlnProLeuHisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSer 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CAGCCGCTGCATCTGGGGATCACCGAAGCGGGTGGCGCGCGCAGCGGGGCAGTAAAATCC 660

Qy        221 AlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GCCATTGGTTTAGGTCTGCTGCTGTCTGAAGGCATCGGCGACACGCTGCGCGTATCACTG 720

Qy        241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GCGGCCGATCCGGTCGAAGAGATCAAAGTCGGTTTCGATATTTTGAAATCGCTTCGTATC 780

Qy        261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspVal 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CGTTCGCGAGGTATCAACTTCATCGCCTGTCCGACCTGTTCGCGTCAGGAATTTGATGTT 840

Qy        281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 ATCGGTACAGTTAACGCGCTGGAGCAACGCCTGGAAGATATCATCACTCCGATGGACGTT 900

Qy        301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TCGATTATCGGCTGCGTGGTGAATGGCCCGGGTGAGGCGCTGGTTTCTACACTCGGCGTC 960

Qy        321 ThrGlyGlyAsnLysLysSerGlyLeuTyrGluAspGlyValArgLysAspArgLeuAsp 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ACCGGCGGCAACAAGAAAAGCGGTCTCTATGAAGATGGCGTGCGCAAAGACCGTCTGGAC 1020

Qy        341 AsnAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AACAACGATATGATCGACCAGCTGGAGGCACGCATTCGTGCGAAAGCCAGTCAGCTGGAC 1080

Qy        361 GluAlaArgArgIleAspValGlnGlnValGluLys 372
              ||||||||||||||||||||||||||||||||||||
Db       1081 GAAGCGCGTCGAATTGACGTTCAGCAGGTTGAAAAA 1116

RESULT 2
AX036302
LOCUS       AX036302               1119 bp    DNA     linear   PAT 16-NOV-2000
DEFINITION  Sequence 29 from Patent EP1043403.
ACCESSION   AX036302
```

```
VERSION     AX036302.1  GI:11225912
KEYWORDS    .
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   .
  TITLE     Novel method for identifying antibacterial compounds
  JOURNAL   Patent: EP 1043403-A 29 11-OCT-2000;
            GPC AG GENOME PHARMACEUTICALS (DE)
FEATURES             Location/Qualifiers
     source          1. .1119
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:562"
BASE COUNT      272 a    279 c    317 g    251 t
ORIGIN

Alignment Scores:
Pred. No.:        .          9.1e-136       Length:          1119
Score:                       1866.00        Matches:         372
Percent Similarity:          100.00%        Conservative:    0
Best Local Similarity:       100.00%        Mismatches:      0
Query Match:                 100.00%        Indels:          0
DB:                          6              Gaps:            0

US-09-921-992-78 (1-372) x AX036302 (1-1119)

Qy          1 MetHisAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCATAACCAGGCTCCAATTCAACGTAGAAAATCAACACGTATTTACGTTGGGAATGTG 60

Qy         21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrThrAsp 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCGATTGGCGATGGTGCTCCCATCGCCGTACAGTCCATGACCAATACGCGTACGACAGAC 120

Qy         41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GTCGAAGCAACGGTCAATCAAATCAAGGCGCTGGAACGCGTTGGCGCTGATATCGTCCGT 180

Qy         61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsn 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GTATCCGTACCGACGATGGACGCGGCAGAAGCGTTCAAACTCATCAAACAGCAGGTTAAC 240

Qy         81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTGCCGCTGGTGGCTGACATCCACTTCGACTATCGCATTGCGCTGAAAGTAGCGGAATAC 300

Qy        101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMet 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGCGTCGATTGTCTGCGTATTAACCCTGGCAATATCGGTAATGAAGAGCGTATTCGCATG 360

Qy        121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GTGGTTGACTGTGCGCGCGATAAAAACATTCCGATCCGTATTGGCGTTAACGCCGGATCG 420

Qy        141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSer 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTGGAAAAAGATCTGCAAGAAAAGTATGGCGAACCGACGCCGCAGGCGTTGCTGGAATCT 480

Qy        161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GCCATGCGTCATGTTGATCATCTCGATCGCCTGAACTTCGATCAGTTCAAAGTCAGCGTG 540

Qy        181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AAAGCGTCTGACGTCTTCCTCGCTGTTGAGTCTTATCGTTTGCTGGCAAAACAGATCGAT 600

Qy        201 GlnProLeuHisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSer 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CAGCCGTTGCATCTGGGGATCACCGAAGCCGGTGGTGCGCGCAGCGGGGCAGTAAAATCC 660

Qy        221 AlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GCCATTGGTTTAGGTCTGCTGCTGTCTGAAGGCATCGGCGACACGCTGCGCGTATCGCTG 720

Qy        241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GCGGCCGATCCGGTCGAAGAGATCAAAGTCGGTTTCGATATTTTGAAATCGCTGCGTATC 780

Qy        261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspVal 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CGTTCGCGAGGGATCAACTTCATCGCCTGCCCGACCTGTTCGCGTCAGGAATTTGATGTT 840

Qy        281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 ATCGGTACGGTTAACGCGCTGGAGCAACGCCTGGAAGATATCATCACTCCGATGGACGTT 900

Qy        301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TCGATTATCGGCTGCGTGGTGAATGGCCCAGGTGAGGCGCTGGTTTCTACACTCGGCGTC 960

Qy        321 ThrGlyGlyAsnLysLysSerGlyLeuTyrGluAspGlyValArgLysAspArgLeuAsp 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ACCGGCGGCAACAAGAAAAGCGGCCTCTATGAAGATGGCGTGCGCAAAGACCGTCTGGAC 1020

Qy        341 AsnAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AACAACGATATGATCGACCAGCTGGAAGCACGCATTCGTGCGAAAGCCAGTCAGCTGGAC 1080

Qy        361 GluAlaArgArgIleAspValGlnGlnValGluLys 372
              ||||||||||||||||||||||||||||||||||||
Db       1081 GAAGCGCGTCGAATTGACGTTCAGCAGGTTGAAAAA 1116

RESULT 3
AX038912
LOCUS       AX038912               1119 bp    DNA     linear   PAT 16-NOV-2000
DEFINITION  Sequence 29 from Patent WO0061793.
ACCESSION   AX038912
VERSION     AX038912.1  GI:11228221
KEYWORDS    .
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Loferer,H. and Jacobi,A.
  TITLE     Novel method for identifying antibacterial compounds
  JOURNAL   Patent: WO 0061793-A 29 19-OCT-2000;
            GPC BIOTECH AG (DE) ; LOFERER HANNES (DE) ; JACOBI ALEXANDER (DE)
FEATURES             Location/Qualifiers
     source          1. .1119
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:562"
BASE COUNT      272 a    279 c    317 g    251 t
ORIGIN

Alignment Scores:
Pred. No.:                   9.1e-136       Length:          1119
Score:                       1866.00        Matches:         372
Percent Similarity:          100.00%        Conservative:    0
Best Local Similarity:       100.00%        Mismatches:      0
Query Match:                 100.00%        Indels:          0
DB:                          6              Gaps:            0

US-09-921-992-78 (1-372) x AX038912 (1-1119)

Qy          1 MetHisAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCATAACCAGGCTCCAATTCAACGTAGAAAATCAACACGTATTTACGTTGGGAATGTG 60

Qy         21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrThrAsp 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCGATTGGCGATGGTGCTCCCATCGCCGTACAGTCCATGACCAATACGCGTACGACAGAC 120
```

```
Qy        41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 GTCGAAGCAACGGTCAATCAAATCAAGGCGCTGGAACGCGTTGGCGCTGATATCGTCCGT 180

Qy        61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsn 80
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 GTATCCGTACCGACGATGGACGCGGCAGAAGCGTTCAAACTCATCAAACAGCAGGTTAAC 240

Qy        81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 GTGCCGCTGGTGGCTGACATCCACTTCGACTATCGCATTGCGCTGAAAGTAGCGGAATAC 300

Qy       101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMet 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 GGCGTCGATTGTCTGCGTATTAACCCTGGCAATATCGGTAATGAAGAGCGTATTCGCATG 360

Qy       121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 GTGGTTGACTGTGCGCGCGATAAAAACATTCCGATCCGTATTGGCGTTAACGCCGGATCG 420

Qy       141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSer 160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 CTGGAAAAAGATCTGCAAGAAAAGTATGGCGAACCGACGCCGCAGGCGTTGCTGGAATCT 480

Qy       161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 GCCATGCGTCATGTTGATCATCTCGATCGCCTGAACTTCGATCAGTTCAAAGTCAGCGTG 540

Qy       181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 AAAGCGTCTGACGTCTTCCTCGCTGTTGAGTCTTATCGTTTGCTGGCAAAACAGATCGAT 600

Qy       201 GlnProLeuHisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSer 220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601 CAGCCGTTGCATCTGGGGATCACCGAAGCCGGTGGTGCGCGCAGCGGGGCAGTAAAATCC 660

Qy       221 AlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       661 GCCATTGGTTTAGGTCTGCTGCTGTCTGAAGGCATCGGCGACACGCTGCGCGTATCGCTG 720

Qy       241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       721 GCGGCCGATCCGGTCGAAGAGATCAAAGTCGGTTTCGATATTTTGAAATCGCTGCGTATC 780

Qy       261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspVal 280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       781 CGTTCGCGAGGGATCAACTTCATCGCCTGCCCGACCTGTTCGCGTCAGGAATTTGATGTT 840

Qy       281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       841 ATCGGTACGGTTAACGCGCTGGAGCAACGCCTGGAAGATATCATCACTCCGATGGACGTT 900

Qy       301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       901 TCGATTATCGGCTGCGTGGTGAATGGCCCAGGTGAGGCGCTGGTTTCTACACTCGGCGTC 960

Qy       321 ThrGlyGlyAsnLysLysSerGlyLeuTyrGluAspGlyValArgLysAspArgLeuAsp 340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       961 ACCGGCGGCAACAAGAAAAGCGGCCTCTATGAAGATGGCGTGCGCAAAGACCGTCTGGAC 1020

Qy       341 AsnAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1021 AACAACGATATGATCGACCAGCTGGAAGCACGCATTCGTGCGAAAGCCAGTCAGCTGGAC 1080

Qy       361 GluAlaArgArgIleAspValGlnGlnValGluLys 372
             ||||||||||||||||||||||||||||||||||||
Db      1081 GAAGCGCGTCGAATTGACGTTCAGCAGGTTGAAAAA 1116

RESULT 4
AX050487
LOCUS       AX050487               1119 bp    DNA     linear   PAT 12-JAN-2001
DEFINITION  Sequence 1 from Patent WO0072022.
ACCESSION   AX050487
VERSION     AX050487.1  GI:12226691
KEYWORDS    .
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Jomaa,H.
  TITLE     Use of genes of the deoxy-d-xylulose phosphate biosynthetic pathway
            for altering the concentration of isoprenoid
  JOURNAL   Patent: WO 0072022-A 1 30-NOV-2000;
            Jomaa, Hassan (DE)
FEATURES             Location/Qualifiers
     source          1. .1119
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:562"
     CDS             1. .1119
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAC21801.1"
                     /db_xref="GI:12226692"
                     /translation="MHNQAPIQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEAT
                     VNQIKALERVGADIVRVSVPTMDAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGV
                     DCLRINPGNIGNEERIRMVVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLES
                     AMRHVDHLDRLNFDQFKVSVKASDVFLAVESYRLLAKQIDQPLHLGITEAGGARSGAV
                     KSAIGLGLLLSEGIGDTLRVSLAADPVEEIKVGFDILKSLRIRSRGINFIACPTCSRQ
                     EFDVIGTVNALEQRLEDIITPMDVSIIGCVVNGPGEALVSTLGVTGGNKKSGLYEDGV
                     RKDRLDNNDMIDQLEARIRAKASQLDEARRIDVQQVEK"
BASE COUNT      272 a    279 c    317 g    251 t
ORIGIN

Alignment Scores:
Pred. No.:                9.1e-136        Length:          1119
Score:                    1866.00         Matches:         372
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       6               Gaps:            0

US-09-921-992-78 (1-372) x AX050487 (1-1119)

Qy         1 MetHisAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnVal 20
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1 ATGCATAACCAGGCTCCAATTCAACGTAGAAAATCAACACGTATTTACGTTGGGAATGTG 60

Qy        21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrThrAsp 40
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 CCGATTGGCGATGGTGCTCCCATCGCCGTACAGTCCATGACCAATACGCGTACGACAGAC 120

Qy        41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 GTCGAAGCAACGGTCAATCAAATCAAGGCGCTGGAACGCGTTGGCGCTGATATCGTCCGT 180

Qy        61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsn 80
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 GTATCCGTACCGACGATGGACGCGGCAGAAGCGTTCAAACTCATCAAACAGCAGGTTAAC 240

Qy        81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 GTGCCGCTGGTGGCTGACATCCACTTCGACTATCGCATTGCGCTGAAAGTAGCGGAATAC 300

Qy       101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMet 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 GGCGTCGATTGTCTGCGTATTAACCCTGGCAATATCGGTAATGAAGAGCGTATTCGCATG 360

Qy       121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 GTGGTTGACTGTGCGCGCGATAAAAACATTCCGATCCGTATTGGCGTTAACGCCGGATCG 420

Qy       141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSer 160
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          421 CTGGAAAAAGATCTGCAAGAAAAGTATGGCGAACCGACGCCGCAGGCGTTGCTGGAATCT 480

Qy          161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          481 GCCATGCGTCATGTTGATCATCTCGATCGCCTGAACTTCGATCAGTTCAAAGTCAGCGTG 540

Qy          181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          541 AAAGCGTCTGACGTCTTCCTCGCTGTTGAGTCTTATCGTTTGCTGGCAAAACAGATCGAT 600

Qy          201 GlnProLeuHisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSer 220
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          601 CAGCCGTTGCATCTGGGGATCACCGAAGCCGGTGGTGCGCGCAGCGGGGCAGTAAAATCC 660

Qy          221 AlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          661 GCCATTGGTTTAGGTCTGCTGCTGTCTGAAGGCATCGGCGACACGCTGCGCGTATCGCTG 720

Qy          241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          721 GCGGCCGATCCGGTCGAAGAGATCAAAGTCGGTTTCGATATTTTGAAATCGCTGCGTATC 780

Qy          261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspVal 280
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          781 CGTTCGCGAGGGATCAACTTCATCGCCTGCCCGACCTGTTCGCGTCAGGAATTTGATGTT 840

Qy          281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          841 ATCGGTACGGTTAACGCGCTGGAGCAACGCCTGGAAGATATCATCACTCCGATGGACGTT 900

Qy          301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          901 TCGATTATCGGCTGCGTGGTGAATGGCCCAGGTGAGGCGCTGGTTTCTACACTCGGCGTC 960

Qy          321 ThrGlyGlyAsnLysLysSerGlyLeuTyrGluAspGlyValArgLysAspArgLeuAsp 340
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          961 ACCGGCGGCAACAAGAAAAGCGGCCTCTATGAAGATGGCGTGCGCAAAGACCGTCTGGAC 1020

Qy          341 AsnAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1021 AACAACGATATGATCGACCAGCTGGAAGCACGCATTCGTGCGAAAGCCAGTCAGCTGGAC 1080

Qy          361 GluAlaArgArgIleAspValGlnGlnValGluLys 372
                ||||||||||||||||||||||||||||||||||||
Db         1081 GAAGCGCGTCGAATTGACGTTCAGCAGGTTGAAAAA 1116

RESULT 5
AX393943
LOCUS       AX393943                1119 bp    DNA     linear   PAT 23-MAR-2002
DEFINITION  Sequence 3 from Patent WO0212478.
ACCESSION   AX393943
VERSION     AX393943.1  GI:19701905
KEYWORDS    .
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Boronat,A., Campos,N., Rodriguez-Concepcion,M., Rohmer,M.,
            Seeman,M., Valentin,H.E., Venkatesh,T.V. and Venkatramesh,M.
  TITLE     Methyl-d-erythritol phosphate pathway genes
  JOURNAL   Patent: WO 0212478-A 3 14-FEB-2002;
            Monsanto Technology LLC (US)
FEATURES             Location/Qualifiers
     source          1..1119
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:562"
     CDS             1..1119
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAD29065.1"
                     /db_xref="GI:19701906"
                     /translation="MHNQAPIQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEAT
                     VNQIKALERVGADIVRVSVPTMDAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGV
                     DCLRINPGNIGNEERIRMVVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLES
                     AMRHVDHLDRLNFDQFKVSVKASDVFLAVESYRLLAKQIDQPLHLGITEAGGARSGAV
                     KSAIGLGLLLSEGIGDTLRVSLAADPVEEIKVGFDILKSLRIRSRGINFIACPTCSRQ
                     EFDVIGTVNALEQRLEDIITPMDVSIIGCVVNGPGEALVSTLGVTGGNKKSGLYEDGV
                     RKDRLDNNDMIDQLEARIRAKASQLDEARRIDVQQVEK"
BASE COUNT      272 a    279 c    317 g    251 t
ORIGIN

Alignment Scores:
Pred. No.:              9.1e-136       Length:         1119
Score:                  1866.00        Matches:        372
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            100.00%        Indels:         0
DB:                     6              Gaps:           0

US-09-921-992-78 (1-372) x AX393943 (1-1119)

Qy            1 MetHisAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnVal 20
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db            1 ATGCATAACCAGGCTCCAATTCAACGTAGAAAATCAACACGTATTTACGTTGGGAATGTG 60

Qy           21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrThrAsp 40
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           61 CCGATTGGCGATGGTGCTCCCATCGCCGTACAGTCCATGACCAATACGCGTACGACAGAC 120

Qy           41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          121 GTCGAAGCAACGGTCAATCAAATCAAGGCGCTGGAACGCGTTGGCGCTGATATCGTCCGT 180

Qy           61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsn 80
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          181 GTATCCGTACCGACGATGGACGCGGCAGAAGCGTTCAAACTCATCAAACAGCAGGTTAAC 240

Qy           81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          241 GTGCCGCTGGTGGCTGACATCCACTTCGACTATCGCATTGCGCTGAAAGTAGCGGAATAC 300

Qy          101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMet 120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          301 GGCGTCGATTGTCTGCGTATTAACCCTGGCAATATCGGTAATGAAGAGCGTATTCGCATG 360

Qy          121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          361 GTGGTTGACTGTGCGCGCGATAAAAACATTCCGATCCGTATTGGCGTTAACGCCGGATCG 420

Qy          141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSer 160
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          421 CTGGAAAAAGATCTGCAAGAAAAGTATGGCGAACCGACGCCGCAGGCGTTGCTGGAATCT 480

Qy          161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          481 GCCATGCGTCATGTTGATCATCTCGATCGCCTGAACTTCGATCAGTTCAAAGTCAGCGTG 540

Qy          181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          541 AAAGCGTCTGACGTCTTCCTCGCTGTTGAGTCTTATCGTTTGCTGGCAAAACAGATCGAT 600

Qy          201 GlnProLeuHisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSer 220
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          601 CAGCCGTTGCATCTGGGGATCACCGAAGCCGGTGGTGCGCGCAGCGGGGCAGTAAAATCC 660

Qy          221 AlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          661 GCCATTGGTTTAGGTCTGCTGCTGTCTGAAGGCATCGGCGACACGCTGCGCGTATCGCTG 720

Qy          241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          721 GCGGCCGATCCGGTCGAAGAGATCAAAGTCGGTTTCGATATTTTGAAATCGCTGCGTATC 780
```

```
Qy        261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspVal 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CGTTCGCGAGGGATCAACTTCATCGCCTGCCCGACCTGTTCGCGTCAGGAATTTGATGTT 840

Qy        281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 ATCGGTACGGTTAACGCGCTGGAGCAACGCCTGGAAGATATCATCACTCCGATGGACGTT 900

Qy        301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TCGATTATCGGCTGCGTGGTGAATGGCCCAGGTGAGGCGCTGGTTTCTACACTCGGCGTC 960

Qy        321 ThrGlyGlyAsnLysLysSerGlyLeuTyrGluAspGlyValArgLysAspArgLeuAsp 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ACCGGCGGCAACAAGAAAAGCGGCCTCTATGAAGATGGCGTGCGCAAAGACCGTCTGGAC 1020

Qy        341 AsnAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AACAACGATATGATCGACCAGCTGGAAGCACGCATTCGTGCGAAAGCCAGTCAGCTGGAC 1080

Qy        361 GluAlaArgArgIleAspValGlnGlnValGluLys 372
              ||||||||||||||||||||||||||||||||||||
Db       1081 GAAGCGCGTCGAATTGACGTTCAGCAGGTTGAAAAA 1116

RESULT 6
ECGCPE
LOCUS       ECGCPE                  1697 bp    DNA     linear   BCT 08-DEC-1993
DEFINITION  E.coli gcpE gene.
ACCESSION   X64451 S43432
VERSION     X64451.1  GI:41540
KEYWORDS    gcpE gene.
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 1697)
  AUTHORS   Parker,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-FEB-1992) J. Parker, Southern Illinois Univ, Dept of
            Microbiology, S I U, Carbondale IL 62901, USA
REFERENCE   2  (bases 1 to 1697)
  AUTHORS   Baker,J., Franklin,D.B. and Parker,J.
  TITLE     Sequence and characterization of the gcpE gene of Escherichia coli
  JOURNAL   FEMS Microbiol. Lett. 94, 175-180 (1992)
COMMENT     See also J01629 & M11843.
FEATURES             Location/Qualifiers
     source          1. .1697
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /strain="K12"
                     /db_xref="taxon:562"
                     /sub_clone="pSE411"
                     /cell_line="JF298"
                     /clone_lib="lambda transducing phage dguaB-2"
     CDS             <1. .508
                     /note="ORF"
                     /codon_start=2
                     /transl_table=11
                     /protein_id="CAA45782.1"
                     /db_xref="GI:41541"
                     /db_xref="SWISS-PROT:P27434"
                     /translation="TTTDPATTSTPPASVDTTATNTQTPAVTAPAPAVDPQQNAVVSP
                     SQANVDTAATPAPTAATTPDGAAPLPTDQAGVTTPVADPNALVMNFTADCWLEVTDAT
                     GKKLFSGMQRKDGNLNLTGQAPYKLKIGAPAAVQIQYQGKPVDLSRFIRTNQVARLTL
                     NAEQSPAQ"
     gene            535. .1653
                     /gene="gcpE"
     CDS             535. .1653
                     /gene="gcpE"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA45783.1"
                     /db_xref="GI:41542"
                     /db_xref="SWISS-PROT:P27433"
                     /translation="MHNQAPIQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEAT
                     VNQIKALERVGADIVRVSVPTMDAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGV
                     DCLRINPGNIGNEERIRMVVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLES
                     AMRHVDHLDRLNFDQFKVSVKASDVFLAVESYRLLAKQIDQPLHLGITEAGGARSGAV
                     KSAIGLGLLLSEGIGDTLRVSLAADPVEEIKVGFDILKSLRIRSRGINFIACPTCSRQ
                     EFDVIGTVNALEQRLEDIITPMDVSIIGCVVNGPGEALVSTLGVTGGNKKSGLYEDGV
                     RKDRLDNNDMIDQLEARIRAKASQLDEARRIDVQQVEK"
BASE COUNT      417 a    446 c    469 g    365 t
ORIGIN

Alignment Scores:
Pred. No.:                1.58e-135      Length:          1697
Score:                    1866.00        Matches:         372
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       1              Gaps:            0

US-09-921-992-78 (1-372) x ECGCPE (1-1697)

Qy          1 MetHisAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        535 ATGCATAACCAGGCTCCAATTCAACGTAGAAAATCAACACGTATTTACGTTGGGAATGTG 594

Qy         21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrThrAsp 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        595 CCGATTGGCGATGGTGCTCCCATCGCCGTACAGTCCATGACCAATACGCGTACGACAGAC 654

Qy         41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        655 GTCGAAGCAACGGTCAATCAAATCAAGGCGCTGGAACGCGTTGGCGCTGATATCGTCCGT 714

Qy         61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsn 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        715 GTATCCGTACCGACGATGGACGCGGCAGAAGCGTTCAAACTCATCAAACAGCAGGTTAAC 774

Qy         81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        775 GTGCCGCTGGTGGCTGACATCCACTTCGACTATCGCATTGCGCTGAAAGTAGCGGAATAC 834

Qy        101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMet 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        835 GGCGTCGATTGTCTGCGTATTAACCCTGGCAATATCGGTAATGAAGAGCGTATTCGCATG 894

Qy        121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        895 GTGGTTGACTGTGCGCGCGATAAAAACATTCCGATCCGTATTGGCGTTAACGCCGGATCG 954

Qy        141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSer 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        955 CTGGAAAAAGATCTGCAAGAAAAGTATGGCGAACCGACGCCGCAGGCGTTGCTGGAATCT 1014

Qy        161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1015 GCCATGCGTCATGTTGATCATCTCGATCGCCTGAACTTCGATCAGTTCAAAGTCAGCGTG 1074

Qy        181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1075 AAAGCGTCTGACGTCTTCCTCGCTGTTGAGTCTTATCGTTTGCTGGCAAAACAGATCGAT 1134

Qy        201 GlnProLeuHisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSer 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1135 CAGCCGTTGCATCTGGGGATCACCGAAGCCGGTGGTGCGCGCAGCGGGGCAGTAAAATCC 1194

Qy        221 AlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1195 GCCATTGGTTTAGGTCTGCTGCTGTCTGAAGGCATCGGCGACACGCTGCGCGTATCGCTG 1254

Qy        241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1255 GCGGCCGATCCGGTCGAAGAGATCAAAGTCGGTTTCGATATTTTGAAATCGCTGCGTATC 1314
```

```
Qy        261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspVal 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1315 CGTTCGCGAGGGATCAACTTCATCGCCTGCCCGACCTGTTCGCGTCAGGAATTTGATGTT 1374

Qy        281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1375 ATCGGTACGGTTAACGCGCTGGAGCAACGCCTGGAAGATATCATCACTCCGATGGACGTT 1434

Qy        301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1435 TCGATTATCGGCTGCGTGGTGAATGGCCCAGGTGAGGCGCTGGTTTCTACACTCGGCGTC 1494

Qy        321 ThrGlyGlyAsnLysLysSerGlyLeuTyrGluAspGlyValArgLysAspArgLeuAsp 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1495 ACCGGCGGCAACAAGAAAAGCGGCCTCTATGAAGATGGCGTGCGCAAAGACCGTCTGGAC 1554

Qy        341 AsnAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1555 AACAACGATATGATCGACCAGCTGGAAGCACGCATTCGTGCGAAAGCCAGTCAGCTGGAC 1614

Qy        361 GluAlaArgArgIleAspValGlnGlnValGluLys 372
              ||||||||||||||||||||||||||||||||||
Db       1615 GAAGCGCGTCGAATTGACGTTCAGCAGGTTGAAAAA 1650

RESULT 7
AE005481/c
LOCUS       AE005481               11521 bp    DNA     linear   BCT 21-MAR-2001
DEFINITION  Escherichia coli O157:H7 EDL933 genome, contig 3 of 3, section 100
            of 290.
ACCESSION   AE005481 AE005174
VERSION     AE005481.1  GI:12516908
KEYWORDS    .
SOURCE      Escherichia coli O157:H7 EDL933
  ORGANISM  Escherichia coli O157:H7 EDL933
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 11521)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
  JOURNAL   Nature 409 (6819), 529-533 (2001)
  MEDLINE   21074935
   PUBMED   11206551
REFERENCE   2  (bases 1 to 11521)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-OCT-2000) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .11521
                     /organism="Escherichia coli O157:H7 EDL933"
                     /mol_type="genomic DNA"
                     /strain="EDL933"
                     /serotype="O157:H7"
                     /db_xref="taxon:155864"
                     /note="enterohemorrhagic"
     gene            complement(86. .1204)
                     /gene="gcpE"
                     /note="synonym: Z3778"
     CDS             complement(86. .1204)
                     /gene="gcpE"
                     /function="orf; Unknown function"
                     /note="Residues 1 to 372 of 372 are 100.00 pct identical
                     to residues 1 to 372 of 372 from Escherichia coli K-12
                     Strain MG1655: B2515"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAG57625.1"
                     /db_xref="GI:12516909"
                     /translation="MHNQAPIQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEAT
                     VNQIKALERVGADIVRVSVPTMDAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGV
                     DCLRINPGNIGNEERIRMVVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLES
                     AMRHVDHLDRLNFDQFKVSVKASDVFLAVESYRLLAKQIDQPLHLGITEAGGARSGAV
                     KSAIGLGLLLSEGIGDTLRVSLAADPVEEIKVGFDILKSLRIRSRGINFIACPTCSRQ
                     EFDVIGTVNALEQRLEDIITPMDVSIIGCVVNGPGEALVSTLGVTGGNKKSGLYEDGV
                     RKDRLDNNDMIDQLEARIRAKASQLDEARRIDVQQVEK"
     gene            complement(1231. .2244)
                     /gene="yfgA"
                     /note="synonym: Z3779"
     CDS             complement(1231. .2244)
                     /gene="yfgA"
                     /function="putative membrane; Not classified"
                     /note="Residues 1 to 337 of 337 are 99.10 pct identical to
                     residues 1 to 337 of 337 from Escherichia coli K-12 Strain
                     MG1655: B2516"
                     /codon_start=1
                     /transl_table=11
                     /product="putative membrane protein"
                     /protein_id="AAG57626.1"
                     /db_xref="GI:12516910"
                     /translation="MNTEATHDQNEALTTGARLRNAREQLGLSQQAVAERLCLKVSTV
                     RDIEEDKAPADLASTFLRGYIRSYARLVHIPEEELLPGLEKQAPLRAAKVAPMQSFSL
                     GKRRKKRDGWLMTFTWLVLFVVIGLSGAWWWQDHKAQQEEITTMADQSSXELSSNSEQ
                     GQSVPLNTXTTTDPATTSTPPASVDTTATNTQTPVVTAPAPAVDPQQNAVVSPSQANV
                     DTAATPAPTAATTPPDGAAPLPTDQAGVTTPVADPNALVMNFTADCWLEVTDATGKKLF
                     SGMQRKDGNLNLTGQAPYKLKIGAPAAVQIQYQGKPVDLSRFIRTNQVARLTLNAEQS
                     PAQ"
     gene            complement(2529. .3683)
                     /gene="yfgB"
                     /note="synonym: Z3780"
     CDS             complement(2529. .3683)
                     /gene="yfgB"
                     /function="orf; Unknown function"
                     /note="Residues 1 to 384 of 384 are 99.47 pct identical to
                     residues 1 to 384 of 384 from Escherichia coli K-12 Strain
                     MG1655: B2517"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAG57627.1"
                     /db_xref="GI:12516911"
                     /translation="MSEQLVTPENVTTKDGKINLLDLNRQQMREFFKDLGEKPFRADQ
                     VMKWMYHYCCDNFDEMTDINKVLRGKLKEVAEIRAPEVVEEQRSSDGTIKWAIAVGDQ
                     RVETVYIPEDDRATLCVSSQVGCALECKFCSTAQQGFNRNLRVSEIIGQVWRAAKIVG
                     AAKVTGQRPITNVVMMGMGEPLLNLNNVVPAMEIMLDDFGFGLSKRRVTLSTSGVVPA
                     LDKLGDMIDVALAISLHAPNDEIRDEIVPINKXXNXETFLAAVRRYLEKSNANQGRVT
                     IEYVMLDHVNDGTEHAHQLAELLKDTPCKINLIPWNPFPAAPYGRSSNSRIDRFSKVL
                     MSYGFTTIVRKTRGDDIDAACGQLAGDVIDRTKRTLRKRMQGEAIDIKAV"
     gene            complement(3833. .4264)
                     /gene="ndk"
                     /note="synonym: Z3781"
     CDS             complement(3833. .4264)
                     /gene="ndk"
                     /function="enzyme; Nucleotide biosynthesis: Purine
                     ribonucleotide biosynthesis"
                     /note="Residues 1 to 143 of 143 are 100.00 pct identical
                     to residues 1 to 143 of 143 from Escherichia coli K-12
                     Strain MG1655: B2518"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleoside diphosphate kinase"
                     /protein_id="AAG57628.1"
                     /db_xref="GI:12516912"
                     /translation="MAIERTFSIIKPNAVAKNVIGNIFARFEAAGFKIVGTKMLHLTV
                     EQARGFYAEHDGKPFFDGLVEFMTSGPIVVSVLEGENAVQRHRDLLGATNPANALAGT
```

```
                LRADYADSLTENGTHGSDSVESAAREIAYFFGEGEVCPRTR"
misc_feature    4397. .9255
                /note="O-island #106; Region of the EDL933 chromosome not
                homologous to E. coli K-12 MG1655"
gene            complement(4405. .4983)
                /gene="Z3782"
CDS             complement(4405. .4983)
                /gene="Z3782"
                /function="orf; Unknown function"
                /note="No significant matches"
                /codon_start=1
                /transl_table=11
                /protein_id="AAG57629.1"
                /db_xref="GI:12516913"
                /translation="MLWHKEQGITAIAIDLSQAKAWERALLLLNECLRQYGEEGWSVV
                SPHTSEVNVSRRRLFHVQRADVQMTQIQPGLRRLHYLWPQISDARPMMDPQKCQLCGA
                CWRACEQQVFSLNEGHLQINDALCNGCQNCIAVCFHQAMTVELTILPAKIVNLHANRK
                VCKTCQKSFLTFQQNAQNCLYCQRHRYGMRTP"
gene            complement(5259. .6080)
                /gene="Z3783"
CDS             complement(5259. .6080)
                /gene="Z3783"
                /function="putative enzyme; Not classified"
                /note="Residues 2 to 272 of 273 are 25.26 pct identical to
                residues 6 to 276 of 286 from GenPept 118 :
                gi|5002126|gb|AAD37317.1|AF135170_8 (AF135170) dimethyl
                sulfoxide reductase subunit C [Yersinia pestis]"
                /codon_start=1
                /transl_table=11
                /product="putative dimethyl sulfoxide reductase subunit C"
                /protein_id="AAG57630.1"
                /db_xref="GI:12516914"
                /translation="MHELPLMAFTLLLQASVGMTLWLAFLHRHILITSSARPVMRPPL
                LLAFLAGAVGLLISTLHLGYPLNAMHALSHFSSSWLSREIIFGALYLALLGLSTLLVM
                LRKGGWQLLLMLAAVVGIVDVFCMAQVYIHTSIITWQHYNTLIMFLGTVGILGSALVA
                VFRISGILPQIDALRNGCVLVIALLVLLRLLVQPLWIGDLTANAMQIATLPHAPLAML
                GQLKPILTLSWGISVIGMMFFAVGGCKKNIPAALFGSVMLVGSEVMLRFVFFSIG"
gene            complement(6073. .6510)
                /gene="Z3784"
CDS             complement(6073. .6510)
                /gene="Z3784"
                /function="putative enzyme; Not classified"
                /note="Residues 4 to 134 of 145 are 61.83 pct identical to
                residues 57 to 186 of 205 from GenPept 118 :
                gi|1787122|gb|AAC73981.1| (AE000191) anaerobic dimethyl
                sulfoxide reductase subunit B [Escherichia coli]"
                /codon_start=1
                /transl_table=11
                /product="putative dimethyl sulfoxide reductase subunit B"
                /protein_id="AAG57631.1"
                /db_xref="GI:12516915"
                /translation="MSHNVFAYTLSVSCNHCADPICTKNCPTMAMHKRPGDGIVRVNT
                DKCVGCGYCAWSCPYGAPQMNEQTGXMSKCDFCIDLQAKGEQPICVATCPLGAIKFGP
                IDELREKYGDVSDVKGLPDSSITQPNLVIKPHQGAEKEASHHA"
gene            complement(6699. .9080)
                /gene="Z3785"
CDS             complement(6699. .9080)
                /gene="Z3785"
                /function="putative enzyme; Not classified"
                /note="Residues 25 to 792 of 793 are 43.59 pct identical
                to residues 3 to 784 of 785 from GenPept 118 :
                gi|1787121|gb|AAC73980.1| (AE000191) anaerobic dimethyl
                sulfoxide reductase subunit A [Escherichia coli]"
                /codon_start=1
                /transl_table=11
                /product="putative dimethyl sulfoxide reductase subunit A"
                /protein_id="AAG57632.1"
                /db_xref="GI:12516916"
                /translation="MDTPDEKERITFPQVSRRSFLQATSALITLPFISSTAKAQSPDA
                SPEVTAPVADKVVPTCSTFDCGGKCDIRAHMRDGVVTQITTLPDNELDPQMPIMRACV
                RGRGYRKFVYHPDRLKYPMKRVGKRGEGKFERISWDEATTLIADNLKRITQQYGPASR
                YVHVGTAVXGGTFSGDAMARRLLNLTGGYLEYYHSVSLGNTAAATPYTYGVAASGNSM
                DTLLDTKLVILWGHNPTETIFGHTNYYFQKMKQNGTRFIVVDPRYSDTVSSLADQWIP
                LLPTTDNALMDAMMYVIISENLHDKTFIDTYTLGFDENSMPEGVPANESLVAYLFGAK
                DGIHKTPEWAEKITHVPAQSIRQLARDYATTKPAALIQGWGPQRHICGERTARGSTLL
                ASITGNVGIKGGWAAGYGGSSNRKFCVGPDMPENPVQAKISIMNWMQAADDASKVTPQ
                DGLKGVDKLDSNIRLLFSLAGNYLANQNPDVHQAAKLLEDESKIEFIVLSDLFMTPSA
                KYADVLLPETSFMERWNIGETWGTASYLILSEKPIEPDFERRTDYDWLRDVAKKLGVE
                AEFSQGRDEKQWIEHIWEQTRLAMPDENLPDFATLQKTRRHLFKSAPHIAFEANIRDP
                QNNPFPTPSGKIEIFSKRLFDMQDPEIPALSHYVPAFEGPEDKLTAKYPLQLITWKGK
                NRANSTQYANPWLQEVQTQKLWLNPQDAKQRGISEGDSVKIYNDRGVSIIPVEITPRI

Alignment Scores:
Pred. No.:              1.98e-134       Length:         11521
Score:                  1866.00         Matches:        372
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     1               Gaps:           0

US-09-921-992-78 (1-372) x AE005481 (1-11521)

Qy          1 MetHisAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1204 ATGCATAACCAGGCTCCAATTCAACGTAGAAAATCAACACGTATTTACGTTGGGAATGTG 1145

Qy         21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrThrAsp 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1144 CCGATTGGCGATGGTGCTCCCATCGCCGTACAGTCCATGACCAATACGCGTACGACAGAT 1085

Qy         41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1084 GTCGAAGCAACGGTCAATCAAATCAAGGCGCTGGAACGCGTTGGCGCTGATATCGTCCGT 1025

Qy         61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsn 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1024 GTCTCCGTACCGACGATGGACGCGGCAGAAGCGTTCAAACTCATCAAACAGCAGGTTAAC 965

Qy         81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        964 GTGCCGCTGGTGGCTGACATCCACTTCGACTATCGCATTGCGCTGAAAGTAGCGGAATAC 905

Qy        101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMet 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        904 GGCGTCGATTGTCTGCGTATTAACCCTGGCAATATCGGTAATGAAGAGCGTATTCGCATG 845

Qy        121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        844 GTGGTTGACTGTGCGCGCGATAAAAACATTCCGATCCGTATTGGCGTTAACGCCGGATCG 785

Qy        141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSer 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        784 CTGGAAAAAGATCTGCAAGAAAAGTATGGCGAACCGACGCCGCAGGCGTTGCTGGAATCC 725

Qy        161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        724 GCCATGCGCCATGTTGATCATCTCGATCGCCTGAACTTCGATCAGTTCAAAGTCAGCGTG 665

Qy        181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        664 AAAGCGTCTGACGTCTTCCTCGCTGTTGAGTCTTATCGTTTGCTGGCAAAACAAATCGAT 605

Qy        201 GlnProLeuHisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSer 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        604 CAGCCGCTGCATCTGGGGATCACCGAAGCGGGTGGCGCGCGCAGCGGGGCAGTAAAATCC 545

Qy        221 AlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        544 GCCATTGGTTTAGGTCTGCTGCTGTCTGAAGGCATCGGCGACACGCTGCGCGTATCACTG 485

Qy        241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        484 GCGGCCGATCCGGTCGAAGAGATCAAAGTCGGTTTCGATATTTTGAAATCGCTTCGTATC 425

Qy        261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspVal 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        424 CGTTCGCGAGGTATCAACTTCATCGCCTGTCCGACCTGTTCGCGTCAGGAATTTGATGTT 365

Qy        281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        364 ATCGGTACAGTTAACGCGCTGGAGCAACGCCTGGAAGATATCATCACTCCGATGGACGTT 305

Qy        301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        304 TCGATTATCGGCTGCGTGGTGAATGGCCCGGGTGAGGCGCTGGTTTCTACACTCGGCGTC 245

Qy        321 ThrGlyGlyAsnLysLysSerGlyLeuTyrGluAspGlyValArgLysAspArgLeuAsp 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        244 ACCGGCGGCAACAAGAAAAGCGGTCTCTATGAAGATGGCGTGCGCAAAGACCGTCTGGAC 185

Qy        341 AsnAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        184 AACAACGATATGATCGACCAGCTGGAGGCACGCATTCGTGCGAAAGCCAGTCAGCTGGAC 125

Qy        361 GluAlaArgArgIleAspValGlnGlnValGluLys 372
              ||||||||||||||||||||||||||||||||||||
Db        124 GAAGCGCGTCGAATTGACGTTCAGCAGGTTGAAAAA 89

RESULT 8
AE000338/c
LOCUS       AE000338               13176 bp    DNA     linear   BCT 01-DEC-2000
DEFINITION  Escherichia coli K12 MG1655 section 228 of 400 of the complete
            genome.
ACCESSION   AE000338 U00096
VERSION     AE000338.1  GI:1788862
KEYWORDS    .
SOURCE      Escherichia coli K12
  ORGANISM  Escherichia coli K12
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 13176)
  AUTHORS   Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
            Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
            Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
            Mau,B. and Shao,Y.
  TITLE     The complete genome sequence of Escherichia coli K-12
  JOURNAL   Science 277 (5331), 1453-1474 (1997)
  MEDLINE   97426617
   PUBMED   9278503
REFERENCE   2  (bases 1 to 13176)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   3  (bases 1 to 13176)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   4  (bases 1 to 13176)
  AUTHORS   Plunkett,G. III.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-OCT-1998) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT     This sequence was determined by the E. coli Genome Project at the
            University of Wisconsin-Madison (Frederick R. Blattner, director).
            Supported by NIH grants HG00301 and HG01428 (from the Human Genome
            Project and NCHGR). The entire sequence was independently
            determined from E. coli K12 strain MG1655. Predicted open reading
            frames were determined using GeneMark software, kindly supplied by
            Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
            30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that
            have been correlated with genetic loci are being annotated with CG
            Site Nos., unique ID nos. for the genes in the E. coli Genetic
            Stock Center (CGSC) database at Yale University, kindly supplied by
            Mary Berlyn. A public version of the database is accessible
            (http://cgsc.biology.yale.edu). Annotation of the genome is an
            ongoing task whose goal is to make the genome sequence more useful
            by correlating it with other data. Comments to the authors are
            appreciated. Updated information will be available at the E. coli
            Genome Project's World Wide Web site
            (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
            its annotations are periodically updated; this is version M54. No
            sequence changes. Annotation updates: updated gene identifications
            and products; all new functional assignments courtesy of Monica
            Riley; added promoters, protein binding sites, and repeated
            sequences described in reference 1. The unique numeric identifiers
            beginning with a lowercase 'b' assigned to each gene (protein- or
            RNA-encoding) are now designated as gene synonyms instead of
            labels. This should allow them to be searched for in Entrez as gene
            names.
FEATURES             Location/Qualifiers
     source          1. .13176
                     /organism="Escherichia coli K12"
                     /mol_type="genomic DNA"
                     /strain="K12"
                     /sub_strain="MG1655"
                     /db_xref="taxon:83333"
     promoter        complement(49. .81)
                     /note="factor Sigma70; promoter hisS; documented +1 at
                     2638662"
     gene            complement(86. .1204)
                     /gene="gcpE"
                     /note="synonym: b2515"
     CDS             complement(86. .1204)
                     /gene="gcpE"
                     /function="orf; Unknown"
                     /note="f372; 100 pct identical to GCPE_ECOLI SW: P27433"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC75568.1"
                     /db_xref="GI:1788863"
                     /translation="MHNQAPIQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEAT
                     VNQIKALERVGADIVRVSVPTMDAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGV
                     DCLRINPGNIGNEERIRMVVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLES
                     AMRHVDHLDRLNFDQFKVSVKASDVFLAVESYRLLAKQIDQPLHLGITEAGGARSGAV
                     KSAIGLGLLLSEGIGDTLRVSLAADPVEEIKVGFDILKSLRIRSRGINFIACPTCSRQ
                     EFDVIGTVNALEQRLEDIITPMDVSIIGCVVNGPGEALVSTLGVTGGNKKSGLYEDGV
                     RKDRLDNNDMIDQLEARIRAKASQLDEARRIDVQQVEK"
     gene            complement(1231. .2244)
                     /gene="yfgA"
                     /note="synonym: b2516"
     CDS             complement(1231. .2244)
                     /gene="yfgA"
                     /function="putative membrane; Not classified"
                     /note="f337; 100 pct identical to YFGA_ECOLI SW: P27434"
                     /codon_start=1
                     /transl_table=11
                     /product="putative membrane protein"
                     /protein_id="AAC75569.1"
                     /db_xref="GI:1788864"
                     /translation="MNTEATHDQNEALTTGARLRNAREQLGLSQQAVAERLCLKVSTV
                     RDIEEDKAPADLASTFLRGYIRSYARLVHIPEEELLPGLEKQAPLRAAKVAPMQSFSL
                     GKRRKKRDGWLMTFTWLVLFVVIGLSGAWWWQDRKAQQEEITTMADQSSAELSSNSEQ
                     GQSVPLNTSTTTDPATTSTPPASVDTTATNTQTPAVTAPAPAVDPQQNAVVSPSQANV
                     DTAATPAPTAATTPDGAAPLPTDQAGVTTPVADPNALVMNFTADCWLEVTDATGKKLF
                     SGMQRKDGNLNLTGQAPYKLKIGAPAAVQIQYQGKPVDLSRFIRTNQVARLTLNAEQS
                     PAQ"
     promoter        complement(1288. .1316)
                     /gene="yfgA"
                     /note="factor Sigma70; predicted +1 start at 2639901"
     promoter        complement(1360. .1389)
                     /gene="yfgA"
                     /note="factor Sigma70; predicted +1 start at 2639973"
     gene            complement(2529. .3683)
                     /gene="yfgB"
                     /note="synonym: b2517"
```

```
CDS             complement(2529. .3683)
                /gene="yfgB"
                /function="orf; Unknown"
                /note="f384; 100 pct identical to YFGB_ECOLI SW: P36979"
                /codon_start=1
                /transl_table=11
                /product="orf, hypothetical protein"
                /protein_id="AAC75570.1"
                /db_xref="GI:1788865"
                /translation="MSEQLVTPENVTTKDGKINLLDLNRQQMREFFKDLGEKPFRADQ
                VMKWMYHYCCDNFDEMTDINKVLRGKLKEVAEIRAPEVVEEQRSSDGTIKWAIAVGDQ
                RVETVYIPEDDRATLCVSSQVGCALECKFCSTAQQGFNRNLRVSEIIGQVWRAAKIVG
                AAKVTGQRPITNVVMMGMGEPLLNLNNVVPAMEIMLDDFGFGLSKRRVTLSTSGVVPA
                LDKLGDMIDVALAISLHAPNDEIRDEIVPINKKYNIETFLAAVRRYLEKSNANQGRVT
                IEYVMLDHVNDGTEHAHQLAELLKDTPCKINLIPWNPFPGAPYGRSSNSRIDRFSKVL
                MSYGFTTIVRKTRGDDIDAACGQLAGDVIDRTKRTLRKRMQGEAIDIKAV"
promoter        complement(3778. .3806)
                /note="factor Sigma70; predicted +1 start at 2642391"
promoter        complement(3786. .3814)
                /note="factor Sigma70; predicted +1 start at 2642399"
gene            complement(3833. .4264)
                /gene="ndk"
                /note="synonym: b2518"
CDS             complement(3833. .4264)
                /gene="ndk"
                /EC_number="2.7.4.6"
                /function="enzyme; Purine ribonucleotide biosynthesis"
                /note="f143; 100 pct identical to NDK_ECOLI SW: P24233"
                /codon_start=1
                /transl_table=11
                /product="nucleoside diphosphate kinase"
                /protein_id="AAC75571.1"
                /db_xref="GI:1788866"
                /translation="MAIERTFSIIKPNAVAKNVIGNIFARFEAAGFKIVGTKMLHLTV
                EQARGFYAEHDGKPFFDGLVEFMTSGPIVVSVLEGENAVQRHRDLLGATNPANALAGT
                LRADYADSLTENGTHGSDSVESAAREIAYFFGEGEVCPRTR"
promoter        complement(4322. .4350)
                /note="factor Sigma70; predicted +1 start at 2642935"
gene            complement(4413. .6725)
                /gene="pbpC"
                /note="synonym: b2519"
CDS             complement(4413. .6725)
                /gene="pbpC"
                /function="putative enzyme; Not classified"
                /note="f770; This 770 aa ORF is 31 pct identical (43 gaps)
                to 480 residues of an approx. 720 aa protein PBPF_BACSU
                SW: P38050"
                /codon_start=1
                /transl_table=11
                /product="putative peptidoglycan enzyme"
                /protein_id="AAC75572.1"
                /db_xref="GI:1788867"
                /translation="MPRLLTKRGCWITLAAAPFLLFLAAWGADKLWPLPLHEVNPARV
                VVAQDGTPLWRFADADGIWRYPVTIEDVSPRYLEALINYEDRWFWKHPGVNPFSVARA
                AWQDLTSGRVISGGSTLTMQVARLLDPHPKTFGGKIRQLWRALQLEWHLSKREILTLY
                LNRAPFGGTLQGIGAASWAYLGKSPANLSYSEAAMLAVLPQAPSRLRPDRWPERAEAA
                RNKVLERMAVQGVWSREQVKESREEPIWLAPRQMPQLAPLFSRMMLGKSKSDKITTTL
                DAGLQRRLEELAQNWKGRLPPRSSLAMIVVDHTDMRVRGWVGSVDLNDDSRFGHVDMV
                NSIRSPGSVLKPFVYGLALDEGLIHPASLLQDVPRRTGDYRPGNFDSGFHGPISMSEA
                LVRSLNLPAVQVLEAYGPKRFAAKLRNVGLPLYLPNGAAPNLSLILGGAGAKLEDMAA
                AYTAFARHGKAGKLRLQPDDPLLERPLMSSGAAWIIRRIMADEAQPLPDSALPRVAPL
                AWKTGTSYGYRDAWAIGVNARYVIGIWTGRPDGTPVVGQFGFASAVPLLNQVNNILLS
                RSANLPEDPRPNSVTRGVICWPGGQSLPEGDGNCRRRLATWLLDGSQPPTLLLPEQEG
                INGIRFPIWLDENGKRVAADCPQARQEMINVWPLPLEPWLPASERRAVRLPPASTSCP
                PYGHDAQLPLQLTGVRDGAIIKRLPGAAEATLPLQSSGGAGERWWFLNGEPLTERGRN
                VTLHLTDKGDYQLLVMDDVGQIATVKFVMQ"
gene            complement(6726. .11687)
                /gene="b2520"
CDS             complement(6726. .11687)
                /gene="b2520"
                /function="orf; Unknown"
                /note="f1653; f1653"
                /codon_start=1
                /transl_table=11
                /product="orf, hypothetical protein"
                /protein_id="AAC75573.1"
                /db_xref="GI:1788868"
                /translation="MKKLRVAACMLMLALAGCDNNDNAPTAVKKDAPSEVTKAASSEN
```

Alignment Scores:

Pred. No.:	2.36e-134	Length:	13176
Score:	1866.00	Matches:	372
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	1	Gaps:	0

US-09-921-992-78 (1-372) x AE000338 (1-13176)

```
Qy          1 MetHisAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1204 ATGCATAACCAGGCTCCAATTCAACGTAGAAAATCAACACGTATTTACGTTGGGAATGTG 1145

Qy         21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrThrAsp 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1144 CCGATTGGCGATGGTGCTCCCATCGCCGTACAGTCCATGACCAATACGCGTACGACAGAC 1085

Qy         41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1084 GTCGAAGCAACGGTCAATCAAATCAAGGCGCTGGAACGCGTTGGCGCTGATATCGTCCGT 1025

Qy         61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsn 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1024 GTATCCGTACCGACGATGGACGCGGCAGAAGCGTTCAAACTCATCAAACAGCAGGTTAAC 965

Qy         81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        964 GTGCCGCTGGTGGCTGACATCCACTTCGACTATCGCATTGCGCTGAAAGTAGCGGAATAC 905

Qy        101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMet 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        904 GGCGTCGATTGTCTGCGTATTAACCCTGGCAATATCGGTAATGAAGAGCGTATTCGCATG 845

Qy        121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        844 GTGGTTGACTGTGCGCGCGATAAAAACATTCCGATCCGTATTGGCGTTAACGCCGGATCG 785

Qy        141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSer 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        784 CTGGAAAAAGATCTGCAAGAAAAGTATGGCGAACCGACGCCGCAGGCGTTGCTGGAATCT 725

Qy        161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        724 GCCATGCGTCATGTTGATCATCTCGATCGCCTGAACTTCGATCAGTTCAAAGTCAGCGTG 665

Qy        181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        664 AAAGCGTCTGACGTCTTCCTCGCTGTTGAGTCTTATCGTTTGCTGGCAAAACAGATCGAT 605

Qy        201 GlnProLeuHisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSer 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        604 CAGCCGTTGCATCTGGGGATCACCGAAGCCGGTGGTGCGCGCAGCGGGGCAGTAAAATCC 545

Qy        221 AlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        544 GCCATTGGTTTAGGTCTGCTGCTGTCTGAAGGCATCGGCGACACGCTGCGCGTATCGCTG 485

Qy        241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        484 GCGGCCGATCCGGTCGAAGAGATCAAAGTCGGTTTCGATATTTTGAAATCGCTGCGTATC 425

Qy        261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspVal 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        424 CGTTCGCGAGGGATCAACTTCATCGCCTGCCCGACCTGTTCGCGTCAGGAATTTGATGTT 365

Qy        281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        364 ATCGGTACGGTTAACGCGCTGGAGCAACGCCTGGAAGATATCATCACTCCGATGGACGTT 305

Qy        301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        304 TCGATTATCGGCTGCGTGGTGAATGGCCCAGGTGAGGCGCTGGTTTCTACACTCGGCGTC 245

Qy        321 ThrGlyGlyAsnLysLysSerGlyLeuTyrGluAspGlyValArgLysAspArgLeuAsp 340
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        244 ACCGGCGGCAACAAGAAAAGCGGCCTCTATGAAGATGGCGTGCGCAAAGACCGTCTGGAC 185

Qy        341 AsnAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        184 AACAACGATATGATCGACCAGCTGGAAGCACGCATTCGTGCGAAAGCCAGTCAGCTGGAC 125

Qy        361 GluAlaArgArgIleAspValGlnGlnValGluLys 372
                ||||||||||||||||||||||||||||||||||
Db        124 GAAGCGCGTCGAATTGACGTTCAGCAGGTTGAAAAA 89

RESULT 9
AP002561/c
LOCUS       AP002561               296827 bp    DNA     linear   BCT 24-APR-2003
DEFINITION  Escherichia coli O157:H7 DNA, complete genome, section 12/20.
ACCESSION   AP002561 BA000007
VERSION     AP002561.1  GI:13362592
KEYWORDS    .
SOURCE      Escherichia coli O157:H7
  ORGANISM  Escherichia coli O157:H7
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
            Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
            Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
            Sasakawa,C. and Shinagawa,H.
  TITLE     Complete nucleotide sequence of the prophage VT2-Sakai carrying the
            verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
            derived from the Sakai outbreak
  JOURNAL   Genes Genet. Syst. 74 (5), 227-239 (1999)
  MEDLINE   20198780
   PUBMED   10734605
REFERENCE   2
  AUTHORS   Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
            Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and
            Hayashi,T.
  TITLE     Comparative analysis of the whole set of rRNA operons between an
            enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
            Escherichia coli K-12 strain MG1655
  JOURNAL   Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
  MEDLINE   20557356
   PUBMED   11108008
REFERENCE   3
  AUTHORS   Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
            Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
            Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and
            Shinagawa,H.
  TITLE     Complete nucleotide sequence of the prophage VT1-Sakai carrying the
            Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
            O157:H7 strain derived from the Sakai outbreak
  JOURNAL   Gene 258 (1-2), 127-139 (2000)
  MEDLINE   20564182
   PUBMED   11111050
REFERENCE   4
  AUTHORS   Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
            Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T.,
            Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
            Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
            Shinagawa,H.
  TITLE     Complete genome sequence of enterohemorrhagic Escherichia coli
            O157:H7 and genomic comparison with a laboratory strain K-12
  JOURNAL   DNA Res. 8 (1), 11-22 (2001)
  MEDLINE   21156231
   PUBMED   11258796
REFERENCE   5  (bases 1 to 296827)
  AUTHORS   Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
            Hayashi,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
            Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
            Japan (E-mail:ken@gen-info.osaka-u.ac.jp,
            URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365,
            Fax:81-6-6879-2047)
COMMENT     genome project.
FEATURES             Location/Qualifiers
     source          1. .296827
                     /organism="Escherichia coli O157:H7"
                     /mol_type="genomic DNA"
                     /strain="O157:H7"
                     /sub_strain="RIMD 0509952"
                     /db_xref="taxon:83334"
     gene            414. .1484
                     /gene="ECs3123"
     CDS             414. .1484
                     /gene="ECs3123"
                     /note="unknown"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="BAB36546.1"
                     /db_xref="GI:13362593"
                     /translation="MPNTSIHLSRCNILQNNKLQPTEIYKESQQTAKLEFFCDEFLKI
                     SQSRYGLSTSTDSIANLLAFFTKASDAIDSIKTQKIDVHSYGFVPIRHFVEHVITYKN
                     EFAADYEPYTLTFTRGENNEGILSVESKEGFISQRTINLNEYETAINIINEHVIKENI
                     HNTVQSPTEKYISKINSSDKPYKTSSKESIKSQLYSDQRNIQTCYYILRKNTEWYKYA
                     SSEERYDKFKNSSKEIKNTYKQIVQAQKKLNQMKYINKLSGELIDITDKKLAPLINDS
                     FSYTRDFFAYSKQEKNIFTFDNSKFVDPKEKEGLMVQHSNGQLVITGKYCPEGVQTAF
                     TQEQYDKLIRYINIFFTFPKLE"
     gene            complement(1708. .2784)
                     /gene="ECs3124"
     CDS             complement(1708. .2784)
                     /gene="ECs3124"
                     /note="similar to GLPQ_ECOLI gi|1788572 percent identity
                     99 in 358 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="glycerophosphodiester phosphodiesterase"
                     /protein_id="BAB36547.1"
                     /db_xref="GI:13362594"
                     /translation="MKLTLKNLSMAIMMSTIVMGSSAMAADSNEKIVIAHRGASGYLP
                     EHTLPAKAMAYAQGADYLEQDLVMTKDDHLVVLHDHYLDRVTDVADRFPDRARKDGRY
                     YAIDFTLDEIKSLKFTEGFDIENGKKVQTYPGRFPMGKSDFRVHTFEEEIEFVQGLNH
                     STGKNIGIYPEIKAPWFHHQEGKDIAAKTLEVLKKYGYTGKDDKVYLQCFDADELKRI
                     KNELEPKMGMDLNLVQLIAYTDWNETQQKQPDGSWVNYNYDWMFKPGAMKQVAEYADG
                     IGPDYHMLIEETSQPGNIKLTGMVQDAQQNKLVVHPYTVRSDKLPEYTTDVNQLYDAL
                     YNKAGVNGLFTDFPDKAVKFLNKE"
     gene            complement(2789. .4147)
                     /gene="ECs3125"
     CDS             complement(2789. .4147)
                     /gene="ECs3125"
                     /note="similar to GLPT_ECOLI gi|1788573 percent identity
                     99 in 452 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="sn-glycerol-3-phosphate permease"
                     /protein_id="BAB36548.1"
                     /db_xref="GI:13362595"
                     /translation="MLSIFKPAPHKARLPAAEIDPTYRRLRWQIFLGIFFGYAAYYLV
                     RKNFALAMPYLVEQGFSRGDLGFALSGISIAYGFSKFIMGSVSDRSNPRVFLPAGLIL
                     AAAVMLFMGFVPWATSSIAVMFVLLFLCGWFQGMGWPPCGRTMVHWWSQKERGGIVSV
                     WNCAHNVGGGIPPLLFLLGMAWFNDWHAALYMPAFCAILVALFAFAMMRDTPQSCGLP
                     PIEEYKNDYPDDYNEKAEQELTAKQIFMQYVLPNKLLWYIAIANVFVYLLRYGILDWS
                     PTYLKEVKHFALDKSSWAYFFYEYAGIPGTLLCGWMSDKVFRGNRGATGVFFMTLVTI
                     ATIVYWMNPAGNPTVDMICMIVIGFLIYGPVMLIGLHALELAPKKAAGTAAGFTGLFG
```

```
                YLGGSVAASATVGYTVDFFGWDGGFMVMIGGSILAVILLIVVMIGEKRRHQELLQKRN
                GG"
gene            4420. .6048
                /gene="ECs3126"
CDS             4420. .6048
                /gene="ECs3126"
                /note="similar to GLPA_ECOLI gi|1788574 percent identity
                100 in 542 aa (Conserved in E.coli K-12)"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="anaerobic sn-glycerol-3-phosphate dehydrogenase
                large subunit"
                /protein_id="BAB36549.1"
                /db_xref="GI:13362596"
                /translation="MKTRDSQSSDVIIIGGGATGAGIARDCALRGLRVILVERHDIAT
                GATGRNHGLLHSGARYAVTDAESARECISENQILKRIARHCVEPTNGLFITLPEDDLS
                FQATFIRACEEAGISAEAIDPQQARIIEPAVNPALIGAVKVPDGTVDPFRLTAANMLD
                AKEHGAVILTAHEVTGLIREGATVCGVRVRNHLTGETQALHAPVVVNAAGIWGQHIAE
                YADLRIRMFPAKGSLLIMDHRINQHVINRCRKPSDADILVPGDTISLIGTTSLRIDYN
                EIDDNRVTAEEVDILLREGEKLAPVMAKTRILRAYSGVRPLVASDDDPSGRNVSRGIV
                LLDHAERDGLDGFITITGGKLMTYRLMAEWATDAVCRKLGNTRPCTTADLALPGSQEP
                AEVTLRKVISLPAPLRGSAVYRHGDRTPAWLSEGRLHRSLVCECEAVTAGEVQYAVEN
                LNVNSLLDLRRRTRVGMGTCQGELCACRAAGLLQRFNVTTSAQSIEQLSTFLNERWKG
                VQPIAWGDALRESEFTRWVYQGLCGLEKEQKDAL"
gene            6038. .7297
                /gene="ECs3127"
CDS             6038. .7297
                /gene="ECs3127"
                /note="similar to GLPB_ECOLI gi|1788575 percent identity
                98 in 419 aa (Conserved in E.coli K-12)"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="anaerobic sn-glycerol-3-phosphate dehydrogenase
                membrane anchor subunit"
                /protein_id="BAB36550.1"
                /db_xref="GI:13362597"
                /translation="MRFDTIIMGGGLAGLLCGLQLQKHGLRCAIVTRGQSALHFSSGS
                LDLLSHLPDGQAVTDIHSGLESLRQQAPAHPYSLLGPQRVLDLACQAQALIAESGAQL
                QGSVELPHQRITPLGTLRSTWLSSPEVPVWPLPAKKICVVGISGLMDFQAHLAAASLR
                ELDLAVETAEIELPELDVLRNNATEFRAVNIARFLDNEENWPLLLDALIPVANTCEMI
                LMPACFGLADDKLWRWLNEKLPCSLMLLPTLPPSVLGIRLQNQLQRQFVRQGGVWMPG
                DEVKKVTCKNGVVNEIWTRNHADIPLRPRFAVLASGSFFSGGLVAERNGIREPILGLD
                VLQTATRGEWYKGDFFAPQPWQQFGVTTDQTLRPSQAGQTIENLFTIGSVLGGFDPIA
                QGCGGGVCAVSALHAAQQIAQRAGGQQ"
gene            7294. .8484
                /gene="ECs3128"
CDS             7294. .8484
                /gene="ECs3128"
                /note="similar to GLPC_ECOLI gi|1788576 percent identity
                100 in 396 aa (Conserved in E.coli K-12)"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="anaerobic sn-glycerol-3-phosphate dehydrogenase
                K-small subunit"
                /protein_id="BAB36551.1"
                /db_xref="GI:13362598"
                /translation="MNDTSFENCIKCTVCTTACPVSRVNPGYPGPKQAGPDGERLRLK
                DGALYDEAIKYCINCKRCEVACPSDVKIGDIIQRARAKYDTTRPSLRNFVLSHTDLMG
                SVSTPFAPIVNTATSLKPVRQLLDAALKIDHRRTLPKYSFGTFRRWYRSVAAQQAQYK
                DQVAFFHGCFVNYNHPQLGKDLIKVLNAMGTGVQLLSKEKCCGVPLIANGFTDKARKQ
                AITNVESIREAVGVKGIPVIATSSTCTFALRDEYPEVLNVDNKGLRDHIELATRWLWR
                KLDEGKTLPLKPLPLKVVYHTPCHMEKMGWTLYTLELLRNIPGLELTVLDSQCCGIAG
                TYGFKKENYETSQAIGAPLFRQIEESGADLVVTDCETCKWQIEMSTSLRCEHPITLLA
                QALA"
gene            8677. .9603
                /gene="ECs3129"
CDS             8677. .9603
                /gene="ECs3129"
                /note="similar to YFAD_ECOLI gi|1788577 percent identity
                94 in 303 aa but differs at C-ter (Conserved in E.coli
                K-12)"
                /codon_start=1
```

Alignment Scores:

Pred. No.:	1.45e-132	Length:	296827
Score:	1866.00	Matches:	372
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	1	Gaps:	0

US-09-921-992-78 (1-372) x AP002561 (1-296827)

```
Qy          1 MetHisAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     280346 ATGCATAACCAGGCTCCAATTCAACGTAGAAAATCAACACGTATTTACGTTGGGAATGTG 280287

Qy         21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrThrAsp 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     280286 CCGATTGGCGATGGTGCTCCCATCGCCGTACAGTCCATGACCAATACGCGTACGACAGAT 280227

Qy         41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     280226 GTCGAAGCAACGGTCAATCAAATCAAGGCGCTGGAACGCGTTGGCGCTGATATCGTCCGT 280167

Qy         61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsn 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     280166 GTCTCCGTACCGACGATGGACGCGGCAGAAGCGTTCAAACTCATCAAACAGCAGGTTAAC 280107

Qy         81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     280106 GTGCCGCTGGTGGCTGACATCCACTTCGACTATCGCATTGCGCTGAAAGTAGCGGAATAC 280047

Qy        101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMet 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     280046 GGCGTCGATTGTCTGCGTATTAACCCTGGCAATATCGGTAATGAAGAGCGTATTCGCATG 279987

Qy        121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     279986 GTGGTTGACTGTGCGCGCGATAAAAACATTCCGATCCGTATTGGCGTTAACGCCGGATCG 279927

Qy        141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSer 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     279926 CTGGAAAAAGATCTGCAAGAAAAGTATGGCGAACCGACGCCGCAGGCGTTGCTGGAATCC 279867

Qy        161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     279866 GCCATGCGCCATGTTGATCATCTCGATCGCCTGAACTTCGATCAGTTCAAAGTCAGCGTG 279807

Qy        181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     279806 AAAGCGTCTGACGTCTTCCTCGCTGTTGAGTCTTATCGTTTGCTGGCAAAACAAATCGAT 279747

Qy        201 GlnProLeuHisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSer 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     279746 CAGCCGCTGCATCTGGGGATCACCGAAGCGGGTGGCGCGCGCAGCGGGGCAGTAAAATCC 279687

Qy        221 AlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     279686 GCCATTGGTTTAGGTCTGCTGCTGTCTGAAGGCATCGGCGACACGCTGCGCGTATCACTG 279627

Qy        241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     279626 GCGGCCGATCCGGTCGAAGAGATCAAAGTCGGTTTCGATATTTTGAAATCGCTTCGTATC 279567

Qy        261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspVal 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     279566 CGTTCGCGAGGTATCAACTTCATCGCCTGTCCGACCTGTTCGCGTCAGGAATTTGATGTT 279507

Qy        281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     279506 ATCGGTACAGTTAACGCGCTGGAGCAACGCCTGGAAGATATCATCACTCCGATGGACGTT 279447
```

```
Qy        301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     279446 TCGATTATCGGCTGCGTGGTGAATGGCCCGGGTGAGGCGCTGGTTTCTACACTCGGCGTC 279387

Qy        321 ThrGlyGlyAsnLysLysSerGlyLeuTyrGluAspGlyValArgLysAspArgLeuAsp 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     279386 ACCGGCGGCAACAAGAAAAGCGGTCTCTATGAAGATGGCGTGCGCAAAGACCGTCTGGAC 279327

Qy        341 AsnAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     279326 AACAACGATATGATCGACCAGCTGGAGGCACGCATTCGTGCGAAAGCCAGTCAGCTGGAC 279267

Qy        361 GluAlaArgArgIleAspValGlnGlnValGluLys 372
              ||||||||||||||||||||||||||||||||||||
Db     279266 GAAGCGCGTCGAATTGACGTTCAGCAGGTTGAAAAA 279231

RESULT 10
AE016764/c
LOCUS       AE016764               300099 bp    DNA     linear   BCT 24-DEC-2002
DEFINITION  Escherichia coli CFT073 section 10 of 18 of the complete genome.
ACCESSION   AE016764 AE014075
VERSION     AE016764.1  GI:26109116
KEYWORDS    .
SOURCE      Escherichia coli CFT073
  ORGANISM  Escherichia coli CFT073
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 300099)
  AUTHORS   Welch,R.A., Burland,V., Plunkett,G.D. III, Redford,P., Roesch,P.,
            Rasko,D.A., Buckles,E.L., Liou,S.-R., Boutin,A., Hackett,J.,
            Stroud,D., Mayhew,G.F., Rose,D.J., Zhou,S., Schwartz,D.C.,
            Perna,N.T., Mobley,H.L.T., Donnenberg,M.S. and Blattner,F.R.
  TITLE     Extensive Mosiac Structure Revealed by the Complete Genome Sequence
            of Uropathogenic Escherichia coli
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 17020-17024 (2002)
   PUBMED   12471157
REFERENCE   2  (bases 1 to 300099)
  AUTHORS   Welch,R.A., Burland,V., Plunkett,G.D. III, Redford,P., Roesch,P.,
            Rasko,D.A., Buckles,E.L., Liou,S.-R., Boutin,A., Hackett,J.,
            Stroud,D., Mayhew,G.F., Rose,D.J., Zhou,S., Schwartz,D.C.,
            Perna,N.T., Mobley,H.L.T., Donnenberg,M.S. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUN-2002) Genetics Laboratory, University of
            Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .300099
                     /organism="Escherichia coli CFT073"
                     /mol_type="genomic DNA"
                     /strain="CFT073"
                     /db_xref="taxon:199310"
     gene            complement(163. .1383)
                     /gene="fabB"
                     /locus_tag="c2869"
     CDS             complement(163. .1383)
                     /gene="fabB"
                     /locus_tag="c2869"
                     /EC_number="2.3.1.41"
                     /function="enzyme; Fatty acid and phosphatidic acid
                     biosynthesis"
                     /note="Escherichia coli K-12 ortholog: b2323; Escherichia
                     coli O157:H7 ortholog: z3586"
                     /codon_start=1
                     /transl_table=11
                     /product="3-oxoacyl-[acyl-carrier-protein] synthase I"
                     /protein_id="AAN81319.1"
                     /db_xref="GI:26109117"
                     /translation="MKRAVITGLGIVSSIGNNQQEVLASLREGRSGITFSQELKDSGM
                     RSHVWGNVKLDTTGLIDRKVVRFMSDASIYAFLSMEQAIADAGLSPEAYQNNPRVGLI
                     AGSGGGSPRFQVFGADAMRGPRGLKAVGPYVVTKAMASGVSACLATPFKIHGVNYSIS
                     SACATSAHCIGNAVEQIQLGKQDIVFAGGGEELCWEMACEFDAMGALSTKYNDTPEKA
                     SRTYDAHRDGFVIAGGGGMVVVEELEHALARGAHIYAEIVGYGATSDGADMVAPSGEG
                     AVRCMKMAMHGVDTPIDYLNSHGTSTPVGDVKELAAIREVFGDKSPAISATKAMTGHS
                     LGAAGVQEAIYSLLMLEHGFIAPSINIEELDEQAAGLNIVTETTDRELTTVMSNSFGF
                     GGTNATLVMRKLKD"
     gene            1482. .3548
                     /gene="yfcK"
                     /locus_tag="c2870"
     CDS             1482. .3548
                     /gene="yfcK"
                     /locus_tag="c2870"
                     /function="putative"
                     /note="Escherichia coli K-12 ortholog: b2324; Escherichia
                     coli O157:H7 ortholog: z3587"
                     /codon_start=1
                     /transl_table=11
                     /product="Hypothetical protein yfcK"
                     /protein_id="AAN81320.1"
                     /db_xref="GI:26109118"
                     /translation="MRSLTHIAKIEPPPVRRVTYVKHYSIQPANLEFNAEGTPVSRDF
                     DDVYFSNDNGLEETRYVFLGGNHLEARFPEHPHPLFVVAESGFGTGLNFLTLWQAFDQ
                     FREAHPQAQLQRLHFISFEKFPLTRADLALAHQHWPELAPWAEQLQAQWPLPLPGCHR
                     LLLDEGHVTLDLWFGDINELTSQLDDSLNQKVDAWFLDGFAPAKNPDMWTQNLFNAMA
                     RLARPGSTLATFTSAGFVRRGLQEAGFTMQKRKGFGRKREMLCGVMEQTLPFPCSTPW
                     FNRMGSNKQEAAIIGGGIASALLSLALLRRGWQVTLYCADEAPALGASGNRQGALYPL
                     LSKHDEALNRFFSNAFTFARRFYDLLPVKFDHDWCGVTQLGWDEKSQHKIAQMLSMDL
                     PAELAVAVEANAVEQITGVATNCSGITYPQGGWLCPAELTRNVLELAQQQGLQIHYQH
                     QLQDLSRKDDGWLLNFAGDQHATHSVVVLANGHQISGFSQTSSLPVYSVAGQVSHIPT
                     TPELAKLKQVLCYDGYLTPQNPANQHHCIGASYHRGSEETAYSDEDQQQNRQRLIDCF
                     PQVQWAKEVDVSGKEARCGVRCATRDHLPMVGNVPDYDATLVEYASLAEKKDEAVSAP
                     VYDDLFMFAALGSRGLCSAPLCAEILAAQMSEEPIPMDASTLAALNPNRLWVRKLLKG
                     KAVKAG"
     gene            complement(3669. .3947)
                     /gene="yfcL"
                     /locus_tag="c2871"
     CDS             complement(3669. .3947)
                     /gene="yfcL"
                     /locus_tag="c2871"
                     /note="Escherichia coli K-12 ortholog: b2325; Escherichia
                     coli O157:H7 ortholog: z3588"
                     /codon_start=1
                     /transl_table=11
                     /product="Hypothetical protein yfcL"
                     /protein_id="AAN81321.1"
                     /db_xref="GI:26109119"
                     /translation="MIAEFESRILALIDGMVDHASDDELFASGYLRGHLTLAIAELES
                     GDDHSAQAVHTTVSQSLEKAIGAGELSPRDQALVTDMWEHLFQQASQQ"
     gene            complement(3981. .4529)
                     /gene="yfcM"
                     /locus_tag="c2872"
     CDS             complement(3981. .4529)
                     /gene="yfcM"
                     /locus_tag="c2872"
                     /function="putative"
                     /note="Escherichia coli K-12 ortholog: b2326; Escherichia
                     coli O157:H7 ortholog: z3589"
                     /codon_start=1
                     /transl_table=11
                     /product="Hypothetical protein yfcM"
                     /protein_id="AAN81322.1"
                     /db_xref="GI:26109120"
                     /translation="MNSTHHYEQLIEIFNSCFADEFNTRLIKGDDEPIYLPADAEVPY
                     NRIVFAHGFYASAIHEISHWCIAGKARRDLVDFGYWYCPDGRDAQTQSQFEDVEVKPQ
                     ALDWLFCVAAGYPFNVSCDNLEGDFEPDRVVFQRRVHAQVMDYLANGIPERPARFIKA
                     LQNYYHTPELTAEQFPWPEALN"
     gene            complement(4529. .5338)
                     /gene="yfcA"
                     /locus_tag="c2873"
     CDS             complement(4529. .5338)
                     /gene="yfcA"
                     /locus_tag="c2873"
                     /function="putative structure"
                     /note="Escherichia coli K-12 ortholog: b2327; Escherichia
                     coli O157:H7 ortholog: z3590"
                     /codon_start=1
                     /transl_table=11
                     /product="Hypothetical protein yfcA"
```

/protein_id="AAN81323.1"
/db_xref="GI:26109121"
/translation="METFNSLFMVSPLLLGVLFFVAMLAGFIDSIAGGGGLLTIPALM
AAGMSPANALATNKLQACGGSISATIYFIRRKVVSLSDQKLNIAMTFVGSMSGALLVQ
YVQADVLRQILPILVICIGLYFLLMPKLGEEDRQRRMYGLPFALIAGGCVGFYDGFFG
PAAGSFYALAFVTLCGFNLAKATAHAKLLNATSNIGGLLLFILGGKVIWATGFVMLVG
QFLGARMGSRLVLSKGQKLIRPMIVIVSAVMSAKLLYDSHGQEILHWLGMN"
gene complement(5338. .6162)
/gene="mepA"
/locus_tag="c2874"
CDS complement(5338. .6162)
/gene="mepA"
/locus_tag="c2874"
/EC_number="3.4.99.-"
/function="enzyme; Murein sacculus, peptidoglycan"
/note="Escherichia coli K-12 ortholog: b2328; Escherichia
coli O157:H7 ortholog: z3591"
/codon_start=1
/transl_table=11
/product="Penicillin-insensitive murein endopeptidase
precursor"
/protein_id="AAN81324.1"
/db_xref="GI:26109122"
/translation="MNKTAIALLALLASSASLAATPWQKITQPVPGSAQSIGSFSNGC
IVGADTLPIQSEHYQVMRTDQRRYFGHPDLVMFIQRLSRQVSNLGMGTVLIGDMGMPA
GGRFNGGHASHQTGLDVDIFLQLPKTRWTSAQLLRPQALDLVSRDGKHVVPALWKPEI
FSLIKLAAQDKDVTRIFVNPAIKQQLCLDAGTDRDWLRKVRPWFQHRAHMHVRLRCPA
DSLECEDQPLPPPGDGCGAELQSWFEPPKPGTTKPEKKTPPPLPPSCQALLDEHVI"
gene complement(6166. .7254)
/gene="aroC"
/locus_tag="c2875"
CDS complement(6166. .7254)
/gene="aroC"
/locus_tag="c2875"
/EC_number="4.2.3.5"
/function="enzyme; Amino acid biosynthesis: Chorismate"
/note="Escherichia coli K-12 ortholog: b2329; Escherichia
coli O157:H7 ortholog: z3592"
/codon_start=1
/transl_table=11
/product="Chorismate synthase"
/protein_id="AAN81325.1"
/db_xref="GI:26109123"
/translation="MMAGNTIGQLFRVTTFGESHGLALGCIVDGVPPGIPLTEADLQH
DLDRRRPGTSRYTTQRREPDQVKILSGVFEGVTTGTSIGLLIENTDQRSQDYSAIKDV
FRPGHADYTYEQKYGLRDYRGGGRSSARETAMRVAAGAIAKKYLAEKFGIEIRGCLTQ
MGDIPLEIKDWSQVEQNPFFCPDPDKIDALDELMRALKKEGDSIGAKVTVVASGVPAG
LGEPVFDRLDADIAHALMSINAVKGVEIGDGFDVVALRGSQNRDEITKDGFQSNHAGG
ILGGISSGQQIIAHMALKPTSSITVPGRTINRFGEEVEMITKGRHDPCVGIRAVPIAE
AMLAIVLMDHLLRQRAQNADVKTDIPRW"
gene complement(7286. .8551)
/gene="yfcB"
/locus_tag="c2876"
CDS complement(7286. .8551)
/gene="yfcB"
/locus_tag="c2876"
/EC_number="2.1.1.72"
/function="putative"
/note="Escherichia coli K-12 ortholog: b2330; Escherichia
coli O157:H7 ortholog: z3593"
/codon_start=1
/transl_table=11
/product="Hypothetical adenine-specific methylase yfcB"
/protein_id="AAN81326.1"
/db_xref="GI:26109124"
/translation="MAGIGLFLNQPLYWHFADFFTLRSVDNRVLLNLASPRHQLAKQR
LVLLAKCCFLFHLSFSISNFPTVYPTRSTIDRFKRIFPHSPRVNADLSSSSWQTSRRN
YCDHALEEYVDKIFVDEAVNELQTIQDMLRWSVSRFSAANIWYGHGTDNPWDEAVQLV
LPSLYLPLDIPEDMRTARLTSSEKHRIVERVIRRVNERIPVAYLTNRAWFCGHEFYVD
ERVLVPRSPIGELINNKFAGLISKQPQHILDMCTGSGCIAIACAYAFPEAEVDAVDIS
PDALAVAEQNIEEHGLIHNVIPIRSDLFRDLPKVQYDLIVTNPPYVDAEDMSDLPNEY
RHEPELGLASGTDGLKLTRRILGNAADYLADDGVLICEVGNSMVHLMEQYPDVPFTWL

Alignment Scores:

Pred. No.:	1.47e-132	Length:	300099
Score:	1866.00	Matches:	372
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	1	Gaps:	0

US-09-921-992-78 (1-372) x AE016764 (1-300099)

```
Qy          1 MetHisAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     188410 ATGCATAACCAGGCTCCAATTCAACGTAGAAAATCAACACGTATTTACGTTGGGAATGTG 188351

Qy         21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrThrAsp 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     188350 CCGATTGGCGATGGTGCTCCCATCGCCGTACAGTCCATGACCAATACGCGTACGACAGAC 188291

Qy         41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     188290 GTCGAAGCAACGGTCAATCAAATCAAGGCGCTGGAACGTGTTGGCGCTGATATCGTCCGT 188231

Qy         61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsn 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     188230 GTCTCCGTACCGACGATGGACGCGGCAGAAGCGTTCAAACTCATCAAACAGCAGGTTAAC 188171

Qy         81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     188170 GTGCCGCTGGTGGCTGACATCCACTTCGACTATCGCATTGCGCTGAAAGTAGCGGAATAC 188111

Qy        101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMet 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     188110 GGCGTCGATTGTCTGCGTATTAACCCTGGCAATATCGGTAATGAAGAGCGTATTCGCATG 188051

Qy        121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     188050 GTGGTTGACTGTGCGCGCGATAAAAACATTCCGATCCGTATTGGCGTTAATGCCGGATCG 187991

Qy        141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSer 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     187990 TTGGAAAAAGATCTGCAAGAAAAGTATGGCGAACCGACGCCGCAGGCGTTGCTGGAATCT 187931

Qy        161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     187930 GCCATGCGTCATGTTGATCATCTCGATCGCCTGAACTTCGATCAGTTCAAAGTCAGCGTG 187871

Qy        181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     187870 AAAGCGTCTGACGTCTTCCTCGCTGTTGAGTCTTATCGTTTGCTGGCAAAACAAATCGAT 187811

Qy        201 GlnProLeuHisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSer 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     187810 CAGCCGCTGCATCTGGGGATCACCGAAGCCGGTGGTGCGCGTAGCGGGGCAGTAAAATCC 187751

Qy        221 AlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     187750 GCTATTGGTTTAGGTCTGCTGCTGTCTGAAGGCATCGGCGACACGCTGCGCGTATCGCTG 187691

Qy        241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     187690 GCGGCCGATCCGGTCGAAGAGATCAAAGTCGGTTTCGATATTTTGAAATCGCTGCGCATC 187631

Qy        261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspVal 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     187630 CGTTCGCGCGGGATCAACTTCATCGCCTGCCCGACCTGTTCGCGTCAGGAATTTGATGTT 187571

Qy        281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     187570 ATCGGTACGGTTAACGCGCTGGAGCAACGCCTGGAAGATATCATCACTCCGATGGACGTT 187511

Qy        301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     187510 TCGATTATCGGCTGCGTGGTGAATGGCCCAGGTGAGGCGCTGGTTTCTACACTCGGCGTC 187451
```

```
Qy        321 ThrGlyGlyAsnLysLysSerGlyLeuTyrGluAspGlyValArgLysAspArgLeuAsp 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     187450 ACTGGCGGCAACAAGAAAAGCGGCCTCTATGAAGATGGCGTGCGCAAAGACCGTCTGGAC 187391

Qy        341 AsnAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     187390 AACAACGATATGATCGATCAGCTGGAAGCGCGCATTCGTGCGAAAGCCAGTCAGCTGGAC 187331

Qy        361 GluAlaArgArgIleAspValGlnGlnValGluLys 372
              ||||||||||||||||||||||||||||||||||||
Db     187330 GAAGCGCGTCGAATTGACGTTCAGCAGGTTGAAAAA 187295

RESULT 11
AE015271/c
LOCUS       AE015271               10225 bp    DNA     linear   BCT 18-OCT-2002
DEFINITION  Shigella flexneri 2a str. 301 section 234 of 412 of the complete
            genome.
ACCESSION   AE015271 AE005674
VERSION     AE015271.1  GI:24052928
KEYWORDS    .
SOURCE      Shigella flexneri 2a str. 301
  ORGANISM  Shigella flexneri 2a str. 301
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Shigella.
REFERENCE   1  (bases 1 to 10225)
  AUTHORS   Jin,Q., Yuan,Z.H., Xu,J.G., Wang,Y., Shen,Y., Lu,W.C., Wang,J.H.,
            Liu,H., Yang,J., Yang,F., Qu,D., Zhang,X.B., Zhang,J.Y., Yang,G.W.,
            Wu,H.T., Dong,J., Sun,L.L., Xue,Y., Zhao,A.L., Gao,Y.S., Zhu,J.P.,
            Kan,B., Chen,S.X., Yao,Z.J., He,B.K., Chen,R.S., Ma,D.L.,
            Qiang,B.Q., Wen,Y.M., Hou,Y.D. and Yu,J.
  TITLE     Genome sequence of Shigella flexneri 2a: insights into
            pathogenicity through comparison with genomes of Escherichia coli
            K12 and O157
  JOURNAL   Nucleic Acids Res. 30 (20), 4432-4441 (2002)
   PUBMED   12384590
REFERENCE   2  (bases 1 to 10225)
  AUTHORS   Jin,Q., Shen,Y., Wang,J.H., Liu,H., Yang,J., Yang,F., Zhang,X.B.,
            Zhang,J.Y., Yang,G.W., Wu,H.T., Dong,J., Sun,L.L., Xue,Y.,
            Zhao,A.L., Gao,Y.S., Zhu,J.P., Chen,S.X., Yao,Z.J., Wang,Y.,
            Lu,W.C., Qiang,B.Q., Wen,Y.M. and Hou,Y.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAY-2001) Microbial Genome Center of Chinese Ministry
            of Public Health, 100 YingXin Jie, XuanWu Qu, Beijing 100052, P.R.
            China
FEATURES             Location/Qualifiers
     source          1. .10225
                     /organism="Shigella flexneri 2a str. 301"
                     /mol_type="genomic DNA"
                     /strain="301"
                     /serotype="2a"
                     /db_xref="taxon:198214"
     gene            148. .1440
                     /gene="xseA"
                     /locus_tag="SF2555"
     CDS             148. .1440
                     /gene="xseA"
                     /locus_tag="SF2555"
                     /note="Residues 1 to 430 of 430 are 97 pct identical to
                     residues 27 to 456 of a 456 aa protein from Escherichia
                     coli K12 ref: NP_417004.1"
                     /codon_start=1
                     /transl_table=11
                     /product="exonuclease VII, large subunit"
                     /protein_id="AAN44055.1"
                     /db_xref="GI:24052929"
                     /translation="MGQVWISGEISNFTQPASGHWYFTLKDDTAQVRCAMFRNSNRRV
                     TFRPQHGQQVLVRANITLYEPRGDYQIIVESMQPAGEGLLQQKYEQLKAKLQAECLFD
                     QQYKKPLPSPAHCVGVITSKTGAALHDILHVLKRRDPSLPVIIYPTSVQGDDAPGQIV
                     RAIELANQRNECDVLIVGRGGGSLEDLWSFNDERVARAIFASRIPVVSAVGHETDVTI
                     ADFVADLRAPTPSAAAEVVSRNQQELLRQVQSARQRLEMAMDYYLANRTRRFTQIHHR
                     LQQQHPQLRLARQQTMLERLQKRMSFALENQLKRAGQQQQRLTQRLNQQNPQPKIHRT
                     QTRIQQLEYRLAEILRAQLSATRERFGNAVTHLEAVSPLSTLARGYSVTTATDGNVLK
                     KVKQVKTGEMLTTRLEDGWIESEVKNIQPVKKSRKKVH"
     gene            complement(1437. .1688)
                     /gene="yfgJ"
                     /locus_tag="SF2556"
     CDS             complement(1437. .1688)
                     /gene="yfgJ"
                     /locus_tag="SF2556"
                     /note="Residues 1 to 83 of 83 are 97 pct identical to
                     residues 1 to 83 of a 83 aa protein from Escherichia coli
                     K12 ref: NP_417005.1"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, conserved hypothetical protein"
                     /protein_id="AAN44056.1"
                     /db_xref="GI:24052930"
                     /translation="MNVEGMATGGIHMELHCPKCQHVLDQDNGHARCPSCGEFIEMKA
                     LCPDCHQPLQVLKACGAVDYFCQHGHGLISKKRVEFVLA"
     gene            complement(1721. .3232)
                     /gene="yfgK"
                     /locus_tag="SF2557"
     CDS             complement(1721. .3232)
                     /gene="yfgK"
                     /locus_tag="SF2557"
                     /note="Residues 1 to 503 of 503 are 99 pct identical to
                     residues 1 to 503 of a 503 aa protein from Escherichia
                     coli O157:H7 EDL933 ref: NP_289064.1"
                     /codon_start=1
                     /transl_table=11
                     /product="putative GTP-binding factor"
                     /protein_id="AAN44057.1"
                     /db_xref="GI:24052931"
                     /translation="MRCLMIYKNEALNMVPVVALVGRPNVGKSTLFNRLTRTRDALVA
                     DFPGLTRDRKYGRAEIEGREFICIDTGGIDGTEDGVETRMAEQSLLAIEEADVVLFMV
                     DARAGLMPADEAIAKHLRSREKPTFLVVNKTDGLDPDQAVVDFYSLGLGEIYPIAASH
                     GRGVLSLLEHVLLPWMEDLAPQEEVDEDAEYWAQFEAEENGEEEEEDDFDPQSLPIKL
                     AIVGRPNVGKSTLTNRILGEERVVVYDMPGTTRDSIYIPMERDGREYVLIDTAGVRKR
                     GKITDAVEKFSVIKTLQAIEDANVVMLVIDAREGISDQDLSLLGFILNSGRSLVIVVN
                     KWDGLSQEVKEQVKETLDFRLGFIDFARVHFISALHGSGVGNLFESVREAYDSSTRRV
                     GTSMLTRIMTMAVEDHQPPLVRGRRVKLKYAHAGGYNPPIVVIHGNQVKDLPDSYKRY
                     LMNYFRKSLDVMGSPIRIQFKEGENPYANKRNTLTPTQMRKRKRLMKHIMKNK"
     gene            complement(3311. .4489)
                     /gene="yfgL"
                     /locus_tag="SF2558"
     CDS             complement(3311. .4489)
                     /gene="yfgL"
                     /locus_tag="SF2558"
                     /note="Residues 1 to 392 of 392 are 99 pct identical to
                     residues 1 to 392 of a 392 aa protein from Escherichia
                     coli K12 ref: NP_417007.1"
                     /codon_start=1
                     /transl_table=11
                     /product="putative dehydrogenase"
                     /protein_id="AAN44058.1"
                     /db_xref="GI:24052932"
                     /translation="MQLRKLLLPGLLSVTLLSGCSLFNSEEDVVKMSPLPTVENQFTP
                     TTAWSTSVGSGIGNFYSNLHPALADNVVYAADRAGLVKALNADDGKEIWSVSLAEKDG
                     WFSKEPALLSGGVTVSGGHVYIGSEKAQVYALNTSDGTVAWQTKVAGEALSRPVVSDG
                     LVLIHTSNGQLQALNEADGAVKWTVNLDMPSLSLRGESAPATAFGAAVVGGDNGRVSA
                     VLMEQGQMIWQQRISQATGSTEIDRLSDVDTTPVVVNGVVFALAYNGNLTALDLRSGQ
                     IMWKRELGSVNDFIVDGNRIYLVDQNDRVMALTIDGGVTLWAQSDLLHRLLTSPVLYN
                     GNLVVGDSEGYLHWINVEDGRFVAQQKVDSSGFQTEPVAADGKLLIQAKDGTVYSITR
                     "
     gene            complement(4500. .5120)
                     /gene="yfgM"
                     /locus_tag="SF2559"
     CDS             complement(4500. .5120)
                     /gene="yfgM"
                     /locus_tag="SF2559"
                     /note="Residues 1 to 206 of 206 are 98 pct identical to
                     residues 1 to 206 of a 206 aa protein from Escherichia
                     coli K12 ref: NP_417008.1"
                     /codon_start=1
                     /transl_table=11
```

/product="orf, conserved hypothetical protein"
/protein_id="AAN44059.1"
/db_xref="GI:24052933"
/translation="MEIYENENDQVEAVKRFFAENGKALAVGGILGVGALTGWRYWNS
HQVDSARSASLAYQNAVTAVSEGKPDSIPAAEKFAAENKNTYGALASLELAQQFVDKN
ELEKAAAQLQQGLADTSDENLKAVINLRLARVQVQLKQADAALKTLDTIKGEGWAAIV
ADLRGEALLSKGDKQGARSAWEAGVKSDVTPALSEMMQMKINNLSI"

gene complement(5138. .6412)
/gene="hisS"
/locus_tag="SF2560"

CDS complement(5138. .6412)
/gene="hisS"
/locus_tag="SF2560"
/note="Residues 1 to 424 of 424 are 99 pct identical to residues 1 to 424 of a 424 aa protein from Escherichia coli O157:H7 EDL933 ref: NP_289067.1"
/codon_start=1
/transl_table=11
/product="histidine tRNA synthetase"
/protein_id="AAN44060.1"
/db_xref="GI:24052934"
/translation="MAKNIQAIRGMNDYLPGETAIWQRIEGTLKNVLGSYGYSEIRLP
IVEQTPLFKRAIGEVTDVVEKEMYTFEDRNGDSLTLRPEGTAGCVRAGIEHGLLYNQE
QRLWYIGPMFRHERPQKGRYRQFHQLGCEVFGLQGPDIDAELIMLTARWWRALGIFEH
VTLELNSIGSLEARANYRDALVAFLEQHKEKLDEDCKRRMYTNPLRVLDSKNPEVQAL
LNDAPALGDYLDEESREHFAGLCKLLESAGIAYTVNQRLVRGLDYYNRTVFEWVTNSL
GSQGTVCAGGRYDGLVEQLGGRATPAVGFAMGLERLVLLVQAVNPEFKADPVVDIYLV
ASGADTQSAAMALAERLRDELLGVKLMTNHGGGNFKKQFARADKWGARVAVVLGESEV
ANGTAVVKDLRSGEQTAVAQDSVAAHLRTLLG"

gene complement(6515. .7642)
/gene="gcpE"
/locus_tag="SF2561"

CDS complement(6515. .7642)
/gene="gcpE"
/locus_tag="SF2561"
/note="Residues 1 to 372 of 375 are 99 pct identical to residues 1 to 372 of a 372 aa protein from Escherichia coli O157:H7 EDL933 ref: NP_289068.1"
/codon_start=1
/transl_table=11
/product="orf, conserved hypothetical protein"
/protein_id="AAN44061.1"
/db_xref="GI:24052935"
/translation="MHNQAPIQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEAT
VNQIKALERVGADIVRVSVPTMDAAEAFKLIKQRVNVPLVADIHFDYRIALKVAEYGV
DCLRINPGNIGNEERIRMVVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLES
AMRHVDHLDRLNFDQFKVSVKASDVFLAVESYRLLAKQIDQPLHLGITEAGGARSGAV
KSAIGLGLLLSEGIGDTLRVSLAADPVEEIKVGFDILKSLRIRSRGINFIACPTCSRQ
EFDVIGTVNALEQRLEDIITPMDVSIIGCVVNGPGEALVSTLGVTGGNKKSGLYEDGV
RKDRLDNNDMIDQLEARIRAKASQLDEARRIDVQQVGKIIT"

gene complement(7669. .8682)
/gene="yfgA"
/locus_tag="SF2562"

CDS complement(7669. .8682)
/gene="yfgA"
/locus_tag="SF2562"
/note="Residues 1 to 337 of 337 are 99 pct identical to residues 1 to 337 of a 337 aa protein from Escherichia coli O157:H7 ref: NP_311405.1"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="AAN44062.1"
/db_xref="GI:24052936"
/translation="MNTEATHDQNEALTTGARLRNAREQLGLSQQAVAERLCLKVSTV
RDIEEDKAPADLASTFLRGYIRSYARLVHIPEEELLPGLEKQAPLRAAKVAPMQSFSL
GKRRKKRDGWLMTFTWLVLFVVIGLSGAWWWQDHKAQQEEITTMADQSSAELSSNSEQ
GQSVPLNTSTTTDPATTSTPPASVDTTATNTQTPAVTAPAPAVDPQQNAVVSPSQANV
DTAATPVPTAATTPDGAAPLPTDQAGVTTPAADPNALVMNFTADCWLEVTDATGKKLF

Alignment Scores:
Pred. No.: 1.2e-133 Length: 10225
Score: 1855.00 Matches: 370

Percent Similarity: 99.73% Conservative: 1
Best Local Similarity: 99.46% Mismatches: 1
Query Match: 99.41% Indels: 0
DB: 1 Gaps: 0

US-09-921-992-78 (1-372) x AE015271 (1-10225)

```
Qy          1 MetHisAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7642 ATGCATAACCAGGCTCCAATTCAACGTAGAAAATCAACACGTATTTACGTTGGGAATGTG 7583

Qy         21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrThrAsp 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7582 CCGATTGGCGATGGTGCTCCCATCGCCGTACAGTCCATGACCAATACGCGTACGACAGAT 7523

Qy         41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7522 GTCGAAGCAACGGTCAATCAAATCAAGGCGCTGGAACGCGTTGGCGCTGATATCGTCCGT 7463

Qy         61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsn 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db       7462 GTCTCCGTACCGACGATGGACGCGGCAGAAGCGTTCAAACTCATCAAACAGCGGGTTAAC 7403

Qy         81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7402 GTGCCGCTGGTGGCTGACATCCACTTCGACTATCGCATTGCGCTGAAAGTAGCGGAATAC 7343

Qy        101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMet 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7342 GGCGTCGATTGTCTGCGTATTAACCCTGGTAATATCGGTAATGAAGAGCGTATTCGCATG 7283

Qy        121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7282 GTGGTTGACTGTGCGCGCGATAAAAACATTCCGATCCGTATTGGCGTTAACGCCGGATCG 7223

Qy        141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSer 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7222 CTGGAAAAAGATCTGCAAGAAAAGTATGGCGAACCGACGCCGCAGGCGTTGCTGGAATCC 7163

Qy        161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7162 GCCATGCGTCATGTTGATCATCTCGATCGCCTGAACTTCGATCAGTTCAAAGTCAGCGTG 7103

Qy        181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7102 AAAGCGTCTGACGTCTTCCTCGCTGTTGAGTCTTATCGTTTGCTGGCAAAACAGATCGAT 7043

Qy        201 GlnProLeuHisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSer 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7042 CAGCCGCTGCATCTGGGGATCACCGAAGCCGGTGGTGCACGCAGCGGGGCAGTAAAATCC 6983

Qy        221 AlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6982 GCCATTGGTTTAGGTCTGCTGCTGTCGGAAGGCATCGGCGACACGCTGCGTGTATCGCTG 6923

Qy        241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6922 GCGGCCGATCCGGTCGAAGAGATCAAAGTCGGTTTCGATATTTTGAAATCGCTGCGTATC 6863

Qy        261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspVal 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6862 CGTTCGCGAGGGATTAACTTCATCGCCTGCCCGACCTGTTCGCGTCAGGAATTTGATGTT 6803

Qy        281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6802 ATCGGTACGGTTAACGCGCTGGAGCAACGCCTGGAAGATATCATCACTCCGATGGACGTT 6743

Qy        301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6742 TCGATTATCGGCTGCGTGGTGAATGGCCCAGGTGAGGCGCTGGTTTCTACACTCGGCGTC 6683

Qy        321 ThrGlyGlyAsnLysLysSerGlyLeuTyrGluAspGlyValArgLysAspArgLeuAsp 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       6682 ACCGGCGGCAACAAGAAAAGCGGCCTCTATGAAGATGGCGTGCGCAAAGACCGTCTGGAC 6623

Qy        341 AsnAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6622 AACAACGATATGATCGACCAGCTGGAAGCACGCATTCGTGCGAAAGCCAGTCAGCTGGAC 6563

Qy        361 GluAlaArgArgIleAspValGlnGlnValGluLys 372
              |||||||||||||||||||||||||||||   |||
Db       6562 GAAGCGCGTCGAATTGACGTTCAGCAGGTTGGAAAA 6527

RESULT 12
AE016987/c
LOCUS       AE016987               290380 bp    DNA     linear   BCT 22-APR-2003
DEFINITION  Shigella flexneri 2a str. 2457T section 10 of 16 of the complete
            genome.
ACCESSION   AE016987 AE014073
VERSION     AE016987.1  GI:30042162
KEYWORDS    .
SOURCE      Shigella flexneri 2a str. 2457T
  ORGANISM  Shigella flexneri 2a str. 2457T
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Shigella.
REFERENCE   1  (bases 1 to 290380)
  AUTHORS   Wei,J., Goldberg,M.B., Burland,V., Venkatesan,M.M., Deng,W.,
            Fournier,G., Mayhew,G.F., Plunkett,G. III, Rose,D.J., Darling,A.,
            Mau,B., Perna,N.T., Payne,S.M., Runyen-Janecky,L.J., Zhou,S.,
            Schwartz,D.C. and Blattner,F.R.
  TITLE     Complete Genome Sequence and Comparative Genomics of Shigella
            flexneri Serotype 2a Strain 2457T
  JOURNAL   Infect. Immun. 71 (5), 2775-2786 (2003)
   PUBMED   12704152
REFERENCE   2  (bases 1 to 290380)
  AUTHORS   Wei,J., Goldberg,M.B., Burland,V., Venkatesan,M.M., Deng,W.,
            Fournier,G., Mayhew,G.F., Plunkett,G. III, Rose,D.J., Darling,A.,
            Mau,B., Perna,N.T., Payne,S.M., Runyen-Janecky,L.J., Zhou,S.,
            Schwartz,D.C. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUN-2002) Genetics Laboratory, University of
            Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .290380
                     /organism="Shigella flexneri 2a str. 2457T"
                     /mol_type="genomic DNA"
                     /strain="2457T"
                     /serotype="2a"
                     /db_xref="taxon:198215"
     gene            complement(82. .1209)
                     /gene="gcpE"
                     /locus_tag="S2733"
     CDS             complement(82. .1209)
                     /gene="gcpE"
                     /locus_tag="S2733"
                     /note="residues 1 to 372 of 375 are 96.23 pct identical to
                     residues 1 to 372 of 372 from Escherichia coli K-12 :
                     B2515"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAP17888.1"
                     /db_xref="GI:30042163"
                     /translation="MHNQAPIQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEAT
                     VNQIKALERVGADIVRVSVPTMDAAEAFKLIKQRVNVPLVADIHFDYRIALKVAEYGV
                     DCLRINPGNIGNEERIRMVVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLES
                     AMRHVDHLDRLNFDQFKVSVKASDVFLAVESYRLLAKQIDQPLHLGITEAGGARSGAV
                     KSAIGLGLLLSEGIGDTLRVSLAADPVEEIKVGFDILKSLRIRSRGINFIACPTCSRQ
                     EFDVIGTVNALEQRLEDIITPMDVSIIGCVVNGPGEALVSTLGVTGGNKKSGLYEDGV
                     RKDRLDNNDMIDQLEARIRAKASQLDEARRIDVQQVGKIIT"
     gene            complement(1236. .2249)
                     /gene="yfgA"
                     /locus_tag="S2734"
     CDS             complement(1236. .2249)
                     /gene="yfgA"
                     /locus_tag="S2734"
                     /function="putative membrane; Not classified"
                     /note="residues 1 to 337 of 337 are 80.71 pct identical to
                     residues 1 to 337 of 337 from Escherichia coli K-12 :
                     B2516"
                     /codon_start=1
                     /transl_table=11
                     /product="putative membrane protein"
                     /protein_id="AAP17889.1"
                     /db_xref="GI:30042164"
                     /translation="MNTEATHDQNEALTTGARLRNAREQLGLSQQAVAERLCLKVSTV
                     RDIEEDKAPADLASTFLRGYIRSYARLVHIPEEELLPGLEKQAPLRAAKVAPMQSFSL
                     GKRRKKRDGWLMTFTWLVLFVVIGLSGAWWWQDHKAQQEEITTMADQSSAELSSNSEQ
                     GQSVPLNTSTTTDPATTSTPPASVDTTATNTQTPAVTAPAPAVDPQQNAVVSPSQANV
                     DTAATPVPTAATTPDGAAPLPTDQAGVTTPAADPNALVMNFTADCWLEVTDATGKKLF
                     SGMQRKDGNLNLTGQAPYKLKIGAPAAVQIQYQGKPVDLSRFIRTNQVARLTLNAEQS
                     PAQ"
     gene            complement(2534. .3688)
                     /gene="yfgB"
                     /locus_tag="S2735"
     CDS             complement(2534. .3688)
                     /gene="yfgB"
                     /locus_tag="S2735"
                     /note="residues 1 to 384 of 384 are 99.47 pct identical to
                     residues 1 to 384 of 384 from Escherichia coli K-12 :
                     B2517"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAP17890.1"
                     /db_xref="GI:30042165"
                     /translation="MSEQLVTPENVTTKDGKINLLDLNRQQMREFFKDLGEKTFRADQ
                     VMKWMYHYCCDNFDEMTDINKVLRGKLKEVAEIRAPEVVEEQRSSDGTIKWAIAVGDQ
                     RVETVYIPEDDRATLCVSSQVGCALECKFCSTAQQGFNRNLRVSEIIGQVWRAAKIVG
                     AAKVTGQRPITNVVMMGMGEPLLNLNNVVPAMEIMLDDFGFGLSKRRVTLSTSGVVPA
                     LDKLGDMIDVALAISLHAPNDEIRDEIVPINKKYNIETFLAAVRRYLEKSNANQGRVT
                     IEYVMLDHVNDGTEHAHQLAELLKDTPCKINLIPWNPFPDAPYGRSSNSRIDRFSKVL
                     MSYGFTTIVRKTRGDDIDAACGQLAGDVIDRTKRTLRKRMQGEAIDIKAV"
     gene            complement(3838. .4269)
                     /gene="ndk"
                     /locus_tag="S2736"
     CDS             complement(3838. .4269)
                     /gene="ndk"
                     /locus_tag="S2736"
                     /function="enzyme; Purine ribonucleotide biosynthesis"
                     /note="residues 1 to 143 of 143 are 100.00 pct identical
                     to residues 1 to 143 of 143 from Escherichia coli K-12 :
                     B2518"
                     /codon_start=1
                     /transl_table=11
                     /product="nucleoside diphosphate kinase"
                     /protein_id="AAP17891.1"
                     /db_xref="GI:30042166"
                     /translation="MAIERTFSIIKPNAVAKNVIGNIFARFEAAGFKIVGTKMLHLTV
                     EQARGFYAEHDGKPFFDGLVEFMTSGPIVVSVLEGENAVQRHRDLLGATNPANALAGT
                     LRADYADSLTENGTHGSDSVESAAREIAYFFGEGEVCPRTR"
     gene            complement(4418. .6517)
                     /gene="pbpC"
                     /locus_tag="S2737"
                     /note="truncated"
                     /pseudo
     repeat_region   6514. .7961
                     /note="insertion element"
                     /insertion_seq="ISSfl4"
     gene            complement(6534. .7730)
                     /locus_tag="S2738"
     CDS             complement(6534. .7730)
                     /locus_tag="S2738"
                     /function="IS, phage, Tn; Transposon-related functions"
                     /note="residues 1 to 398 of 398 are 90.95 pct identical to
                     residues 1 to 398 of 398 from GenPept :
                     >gb|AAK18438.1|AF348706_127 (AF348706) putative
                     transposase for ISl10 [Shigella flexneri]"
                     /codon_start=1
```

```
                     /transl_table=11
                     /product="ISSfl4 orf"
                     /protein_id="AAP17892.1"
                     /db_xref="GI:30042167"
                     /translation="MTESSDYESVQVFIGVDVGKDTHHAVAINRSGKRLFDKALPNDE
                     NKLRSLISDLKQHGQILLVVDQPATIGALPVAVARSEGVLVGYLPGLAMRRIADLHAG
                     EAKTDVRDAAIIAEAARTLPHALRTLKLADEQIAELSMLCGFDDDLAAQTTQASNRIR
                     GLLTQIHPAPERVLGPRLEHPAVLDLLQRYPSPEKLASLGEKKLAAQLCKLAPRLGKR
                     LAADIAQALAEQTVVVPGTNAAAVVLPRLALQLITLRKQRDEVALEVEQRVLAHPLYP
                     VLTSMPGVGVRTAARLLTEVACRAFASAAHLAAYAGLAPVTRRSGSSIRGEHPSRRGN
                     KALKRALFLSAFAALRDPLSRAYYTRKMSQGKRHNQALIALARRRCDVLFAMMRDGTF
                     YTPQGS"
     gene            7960. .8442
                     /gene="sseA"
                     /locus_tag="S2740"
                     /note="truncated"
                     /pseudo
     gene            complement(9261. .10037)
                     /gene="sseB"
                     /locus_tag="S2741"
     CDS             complement(9261. .10037)
                     /gene="sseB"
                     /locus_tag="S2741"
                     /function="phenotype; Not classified"
                     /note="residues 1 to 258 of 258 are 94.96 pct identical to
                     residues 4 to 261 of 261 from Escherichia coli K-12 :
                     B2522"
                     /codon_start=1
                     /transl_table=11
                     /product="enhancer of serine sensitivity"
                     /protein_id="AAP17893.1"
                     /db_xref="GI:30042168"
                     /translation="MSETKNELEDLLEKAATEPAHRPAFFRTLLESTVWVPGTAAQGE
                     AVVEDSALDLQHWEKEDGTSVIPFFTSLEALQQAVEDEQAFVVMPVRTLFEMTLGETL
                     FLNAKLPTGKEFMPREISLLIGEEGNPLSSQEILEGGESLILSEVAEPPAQMIDSLTT
                     LFKTIKPVKRAFICSIKENEEAQPNLLIGIEADGDIEEIIQATGSVATDTLPGDEPID
                     ICQVKKGEKGISHFITEHIAPFYERRWGGFLRDFKQNRII"
     repeat_region   10137. .10170
                     /note="REP (repetitive extragenic palindromic) element,
                     contains 1 REP sequence"
     gene            complement(10179. .11462)
                     /gene="pepB"
                     /locus_tag="S2742"
     CDS             complement(10179. .11462)
                     /gene="pepB"
                     /locus_tag="S2742"
                     /function="putative enzyme; Not classified"
                     /note="residues 1 to 427 of 427 are 92.50 pct identical to
                     residues 30 to 456 of 456 from Escherichia coli K-12 :
                     B2523"
                     /codon_start=1
                     /transl_table=11
                     /product="putative peptidase"
                     /protein_id="AAP17894.1"
                     /db_xref="GI:30042169"
                     /translation="MTEAMKITLSTQPADARWGEKATYSINNDGITLHLNGADDLGLI
                     QRAARKIDGLGIKHVQLSGEGWDADRCWAFWQGYKGPKGTRKVEWPDLDDAQRQELDN
                     RLMIIDWVRDTINAPAEELGPSQLAQRAVDLISNVAGDRVTYRITKGEDLREQGYMGL
                     HTVGRGSERSPVLLALDYNPTGDKEAPVYACLVGKGITFDSGGYSIKQTAFMDSMKSD
                     MGGAATVTGALAFAIMRGLNKRVKLFLCCADNLISGNAFKLGDIITYRNGKKVEVMNT
                     DAEGRLVLADGLIDASAQKPEMIIDAATLTGAAKTALGNDYHALFSFDDALAGRLLAS
                     ASQENEPFWRLPLAEFHRSQLPSNFAELNNTGSAAYPAGASTAAGFLSHFVENYQQGW
                     LHIDCSATYRKAPVEQWSAGATGLGVRTIANLLTA"
     gene            complement(11521. .11721)
                     /gene="yfhJ"
                     /locus_tag="S2743"
```

```
Alignment Scores:
Pred. No.:              1e-131          Length:         290380
Score:                  1855.00         Matches:        370
Percent Similarity:     99.73%          Conservative:   1
Best Local Similarity:  99.46%          Mismatches:     1
Query Match:            99.41%          Indels:         0
```

```
DB:                 1           Gaps:          0

US-09-921-992-78 (1-372) x AE016987 (1-290380)

Qy          1 MetHisAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1209 ATGCATAACCAGGCTCCAATTCAACGTAGAAAATCAACACGTATTTACGTTGGGAATGTG 1150

Qy         21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrThrAsp 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1149 CCGATTGGCGATGGTGCTCCCATCGCCGTACAGTCCATGACCAATACGCGTACGACAGAT 1090

Qy         41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1089 GTCGAAGCAACGGTCAATCAAATCAAGGCGCTGGAACGCGTTGGCGCTGATATCGTCCGT 1030

Qy         61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsn 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db       1029 GTCTCCGTACCGACGATGGACGCGGCAGAAGCGTTCAAACTCATCAAACAGCGGGTTAAC 970

Qy         81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        969 GTGCCGCTGGTGGCTGACATCCACTTCGACTATCGCATTGCGCTGAAAGTAGCGGAATAC 910

Qy        101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMet 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        909 GGCGTCGATTGTCTGCGTATTAACCCTGGTAATATCGGTAATGAAGAGCGTATTCGCATG 850

Qy        121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        849 GTGGTTGACTGTGCGCGCGATAAAAACATTCCGATCCGTATTGGCGTTAACGCCGGATCG 790

Qy        141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSer 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        789 CTGGAAAAAGATCTGCAAGAAAAGTATGGCGAACCGACGCCGCAGGCGTTGCTGGAATCC 730

Qy        161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        729 GCCATGCGTCATGTTGATCATCTCGATCGCCTGAACTTCGATCAGTTCAAAGTCAGCGTG 670

Qy        181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        669 AAAGCGTCTGACGTCTTCCTCGCTGTTGAGTCTTATCGTTTGCTGGCAAAACAGATCGAT 610

Qy        201 GlnProLeuHisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSer 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        609 CAGCCGCTGCATCTGGGGATCACCGAAGCCGGTGGTGCACGCAGCGGGGCAGTAAAATCC 550

Qy        221 AlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        549 GCCATTGGTTTAGGTCTGCTGCTGTCGGAAGGCATCGGCGACACGCTGCGTGTATCGCTG 490

Qy        241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        489 GCGGCCGATCCGGTCGAAGAGATCAAAGTCGGTTTCGATATTTTGAAATCGCTGCGTATC 430

Qy        261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspVal 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        429 CGTTCGCGAGGGATTAACTTCATCGCCTGCCCGACCTGTTCGCGTCAGGAATTTGATGTT 370

Qy        281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        369 ATCGGTACGGTTAACGCGCTGGAGCAACGCCTGGAAGATATCATCACTCCGATGGACGTT 310

Qy        301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        309 TCGATTATCGGCTGCGTGGTGAATGGCCCAGGTGAGGCGCTGGTTTCTACACTCGGCGTC 250

Qy        321 ThrGlyGlyAsnLysLysSerGlyLeuTyrGluAspGlyValArgLysAspArgLeuAsp 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        249 ACCGGCGGCAACAAGAAAAGCGGCCTCTATGAAGATGGCGTGCGCAAAGACCGTCTGGAC 190

Qy        341 AsnAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        189 AACAACGATATGATCGACCAGCTGGAAGCACGCATTCGTGCGAAAGCCAGTCAGCTGGAC 130

Qy        361 GluAlaArgArgIleAspValGlnGlnValGluLys 372
                 ||||||||||||||||||||||||||||||||   |||
Db        129 GAAGCGCGTCGAATTGACGTTCAGCAGGTTGGAAAA 94

RESULT 13
AE008814/c
LOCUS       AE008814               23647 bp    DNA     linear   BCT 23-APR-2003
DEFINITION  Salmonella typhimurium LT2, section 118 of 220 of the complete
            genome.
ACCESSION   AE008814 AE006468
VERSION     AE008814.1  GI:16421058
KEYWORDS    .
SOURCE      Salmonella typhimurium LT2
  ORGANISM  Salmonella typhimurium LT2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 23647)
  AUTHORS   McClelland,M., Sanderson,K.E., Spieth,J., Clifton,S.W.,
            Latreille,P., Courtney,L., Porwollik,S., Ali,J., Dante,M., Du,F.,
            Hou,S., Layman,D., Leonard,S., Nguyen,C., Scott,K., Holmes,A.,
            Grewal,N., Mulvaney,E., Ryan,E., Sun,H., Florea,L., Miller,W.,
            Stoneking,T., Nhan,M., Waterston,R. and Wilson,R.K.
  TITLE     Complete genome sequence of Salmonella enterica serovar Typhimurium
            LT2
  JOURNAL   Nature 413 (6858), 852-856 (2001)
  MEDLINE   21534948
   PUBMED   11677609
REFERENCE   2  (bases 1 to 23647)
  AUTHORS   .
  CONSRTM   The Salmonella typhimurium Genome Sequencing Project
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-2001) Genome Sequencing Center, Department of
            Genetics, Washington University School of Medicine, 4444 Forest
            Park Boulevard, St. Louis, MO 63108, USA
COMMENT     COMMENT     Supported by NIH grant 5U 01 AI43283

            Coding sequences below are predicted from manually evaluated
            computer analysis, using similarity information and the programs;
            GLIMMER; http://www.tigr.org/softlab/glimmer/glimmer.html and
            GeneMark; http://opal.biology.gatech.edu/GeneMark/

            EC numbers were kindly provided by Junko Yabuzaki and the Kyoto
            Encyclopedia of Genes and Genomes; http://www.genome.ad.jp/kegg/,
            and Pedro Romero and Peter Karp at EcoCyc;
            http://ecocyc.PangeaSystems.com/ecocyc/

            The analyses of ribosome binding sites and promoter binding sites
            were kindly provided by Heladia Salgado, Julio Collado-Vides and
            ReguonDB;
            http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset

            This sequence was finished as follows unless otherwise noted: all
            regions were double stranded, sequenced with an alternate
            chemistries or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one m13 subclone.
FEATURES             Location/Qualifiers
     source          1. .23647
                     /organism="Salmonella typhimurium LT2"
                     /mol_type="genomic DNA"
                     /strain="LT2; SGSC 1412; ATCC 700720"
                     /db_xref="ATCC:700720"
                     /db_xref="taxon:99287"
                     /note="LT2"
     gene            complement(110. .5711)
                     /gene="ratA"
                     /note="synonym: STM2515"
     CDS             complement(110. .5707)
                     /gene="ratA"
                     /note="RatA (gi|4583529)"
                     /codon_start=1
                     /transl_table=11
                     /product="putative outer membrane protein"
                     /protein_id="AAL21409.1"
                     /db_xref="GI:16421059"
                     /translation="MDRNRQVNKVVHFLLTLLIMFAVSIAPAQALLKGGTWQELNSVT
                     GAVNGTAPLADGAIIPLYQGSTLLDPSKTHDIEFSAMPRDFSADATSTSMRAVNSTDT
                     EGDLFSDPPTIAWENRQPPAMGLVWADAATPDTPLSPQPVPNLTFCAQNLAGRQLVAW
                     AQVEDETNVPALWLFTRTGVPNYATIPLLSPKVALNIKPATDTPVTVSGDHVDSAFEA
                     SKVKVGESITLTISTKACDGTPEGNAPFVIRRKDAENRQGVVNNANPVRVGDTELTTT
                     QTEYRGVTDANGNATVVVTQKEGPGVKTPLVVSSVNFPALTAETAVIFTTITSPDSDK
                     ASMYGHMIESATATLNGITYTFTRPKLAAEASGADKSVVDTNETWALFTWSGADNHCD
                     ILPDAEQLVQMRHEHSTLETWTGWPVAGDAEYWSSTKNQLSGYHAAVHMNSASVVRAA
                     DSDTLLVSCVDKAQPAAHPQITLSPQGPYKAQVGESIDLTMTVVDRDTQKPLPYRYME
                     LFIDPAKNRKGEHQDAWDNLRVTVDSEGMSASSPEHYTGVTDVNGQAHLTLKHNSGLG
                     VETPIRIVMPDDEGGNVELSFSVIFTVVTSPDVDGANMWGHMRGVVDAGNLYKRPLLA
                     VEASHKDGQFSENNEEWATFNSVASATAQCGVGQVPDQSSLAHLYSEHAGGQMESEHG
                     WPTEDYFIAADSDASGTVHVNLENGDSGKFTDTPNYLTCSANEMVAVLDVYFNDDPVI
                     KNADAKVGEQVKMNIHSRNALNGMAIGNTDFTITMANGRRRDGLTTGFTDTSNGEMQF
                     DDVGYVAGQVYHGITDANGDATIILTQKKGVGLLTPLNIAPVDSLISTPVSRSVKFTV
                     ATSPDTPAAKMWGHMADTITVGDWTFERPKLAGEVSNPLRTQDESNETWARVAHSDAV
                     GNPDAGGCAANRLPRIDQLEALYNANSGGKIHSIQGWPTYLNYWSSTYQSATTWKLIA
                     LTNGSEFANSNVSIYASCLASDNPVAASITIEPVNPSQWYDGSDVHAVKVKKGETMQL
                     KVTVKDVSGNPIPEAPFVLTRGDGYDRRGEKYTAQDGDDLQGIVTPVVIDGESLAWTT
                     TKMGSQTGTDGTRIISVTRPDTHGTRTAINATLYENAAVSASIDTIFTVVTSPDVSVA
                     RMWGHMAPSLTAADGAVYQRPLLYAELSSTDNTASKQETNETWAVFHGPASEGANPAR
                     CAAGYYPAVEALDSLYSKYPSRTINTAQGWPVYYSYWSGSNSTSFSSGAKLDFYYAVD
                     LADGSRRSENSATSTAWQYQICATTPLPQATQITLTSPQAMDDAIQAVKAKNSESIPL
                     LITTTDAMGNPVPYATFSLKRDAGKARNPDYNKFVATNGTNMTVTPLTGAQQQFYYAT
                     SVLTGATGADGTLALTLAEPGGIGLKNQLTANLNDTPTATSSLPVVFTVLTSPDSDKA
                     NMYGHMPETFTASNGAEFKRPLVAGEPSSEAHTDTYFETNENWIMVNSFNTGNYGGCP
                     MNQMAAIDDFTALYNDHPSGKVATDIGLPVGKRWWAGDSLLKGSTLYWQYKDLKTGKN
                     YSMSENPGNYYLQLCLTTSRSGLNIALSSDAWNADKSAAVAKKGETIPMTVTVTNDAG
                     QPQAGVAVLLTRDYAYSRGAVDKQYIEPGVIGEPVPFTTSPANMMLAPVAPAGTAVAF
                     NNQNGLSTKWSGFTGDDGKLRFTLTQDKSLGLKTSVTAALANQFDEAASVDAIFTVQT
                     SPDTPYASYWGHMPDTVQVNGVTLRRPYLKAELSAAPRDTWPFNNEFWGTNYYYQSEH
                     VETSLTHLCGSQENIASLDDLKALQSVIGTLQWPTTSSWDYVSQDEGQSNKYYCSFNE
                     TTGQTTCTREKATTSGLGSCRVP"
     RBS             complement(5706. .5711)
                     /gene="ratA"
                     /note="putative RBS for ratA; RegulonDB:STMS1H002622"
     gene            complement(5827. .6793)
                     /gene="sinI"
                     /note="synonym: STM2516"
     CDS             complement(5827. .6786)
                     /gene="sinI"
                     /note="SinI (gi|4583530)"
                     /codon_start=1
                     /transl_table=11
                     /product="putative outer membrane protein"
                     /protein_id="AAL21410.1"
                     /db_xref="GI:16421060"
                     /translation="MQATVKRRLTKVALALVVAGYCAAPAVAANGNLKSGQWQIVSEQ
                     TGTIQGTVPWITRAADKTADTDKDHVTVTIDRGDRKIVTEGDKQFHVGDKVTVNWAIG
                     DTEGDLDTDNTATKATVKWVSFSDQNGSDPKDLGTGDSYEIQAADADRYIGIKITPTT
                     TTGDPAVATELLLKDLSTDAGGGSDDDEIPEGPVVDENVHVVIHEKDSNTNLLKNSGT
                     TLKTNTTYQVLLWSDKNGNGTYDAGENVTDQYDYRWKFVGTSKIAGTGTGGIVNESWN
                     DKDLVIPVTNAEAKTAFEGADGGVTVGSDGVQGFGLSIDYKRK"
     RBS             complement(6788. .6793)
                     /gene="sinI"
                     /note="putative RBS for sinI; RegulonDB:STMS1H002623"
     gene            complement(6844. .9042)
                     /gene="sinH"
                     /note="synonym: STM2517"
     CDS             complement(6844. .9036)
                     /gene="sinH"
                     /note="similar to Escherichia coli intimin and Yersinia
                     pestis invasin proteins; (gi|4583531)"
                     /codon_start=1
                     /transl_table=11
                     /product="SinH"
                     /protein_id="AAL21411.1"
```

```
                     /db_xref="GI:16421061"
                     /translation="MVRRIRILVLLLVSGVSVAASGADNSASANDESHSSLPDLASES
                     AKKEEQENKGKSFKEQGADYFINSATQGFDNLTPEALESQARSYLQNQITSSAQSYLE
                     GVMSPYGKIRTSLSVGEGGDLDGSSLDYFIPWYDNQSTLFFSQISAQRKEDRTIGNFG
                     LGVRQNVGNWLLGGNAFYDYDFTRGHRRLGLGTEAWTDYLKFSGNYYHPLSDWKDSED
                     FDFYEERPARGWDIRMESWLPFYPQLGAKLVYEQYYGDEVALFGTDNLQKDPHAVTLG
                     LEYTPVPLVTVGTDYKAGTGDSNDFSVNATVNYQIGTPLAAQLDPENVKIQHSLMGSR
                     TDFVDRNNFIILEYREKDPLDVTLWLKADATNEHPECVIEDTPEAAVGLEKCKWTVNA
                     LINHHYKIISASWQAKNNAARTLVMPVVKANALTEGNNNSWNLVLPAWVNADTEEQRT
                     ALNTWKVRMTLEDEKGNKQNSGVVEITVQQDRKIELIVDNIADTDRSDHSHEASALAD
                     GEDGVVMDLLITDSFGDSTDRNGNELVDDAMTPVLYDSNDKKVTLAQTPCTTETPCVF
                     IASRDKEAGTVTLSSTLPGTFRWKAKADAYGDSNYVDVTFIGDNLSALNAVIYQVKAA
                     NPVNLIGKEDKHPTVNNAYRFLLWRDKNKDGVFQMSEQLTEEEMALYDYQWEFTGQST
                     NGHTGALANTMNEDLVLPVTNKEAAQKFAANVEDGVQGYGIRVTYSQK"
     RBS             complement(9037. .9042)
                     /gene="sinH"
                     /note="putative RBS for sinH; RegulonDB:STMS1H002624"
     gene            complement(9382. .9615)
                     /gene="yfgJ"
                     /note="synonym: STM2518"
     CDS             complement(9382. .9603)
                     /gene="yfgJ"
                     /note="similar to E. coli orf, hypothetical protein
                     (AAC75563.1); Blastp hit to AAC75563.1 (83 aa), 63%
                     identity in aa 13 - 83"
                     /codon_start=1
                     /transl_table=11
                     /product="putative cytoplasmic protein"
                     /protein_id="AAL21412.1"
                     /db_xref="GI:16421062"
                     /translation="MEITCPVCHHALERNGDTAHCETCAKDFSLQALCPDCRQPLQVL
                     KACGAVDYFCQNGHGLISKKRVNFVISDQ"
     RBS             complement(9610. .9615)
                     /gene="yfgJ"
                     /note="putative RBS for yfgJ; RegulonDB:STMS1H002625"
     gene            complement(9692. .11176)
                     /gene="engA"
                     /note="synonym: STM2519"
     CDS             complement(9692. .11164)
                     /gene="engA"
                     /note="similar to E. coli putative GTP-binding factor
                     (AAC75564.1); Blastp hit to AAC75564.1 (503 aa), 97%
                     identity in aa 14 - 503"
                     /codon_start=1
                     /transl_table=11
                     /product="putative GTP-binding protein"
                     /protein_id="AAL21413.1"
                     /db_xref="GI:16421063"
                     /translation="MVPVVALVGRPNVGKSTLFNRLTRTRDALVADFPGLTRDRKYGR
                     AEVEGREFICIDTGGIDGTEDGVETRMAEQSLLAIEEADVVLFMVDARAGLMPADEAI
                     AKHLRSREKPTFLVANKTDGLDPDQAVVDFYSLGLGEIYPIAASHGRGVLSLLEHVLL
                     PWMDDVAPQEEVDEDAEYWAQFEAEQNGEEAPEDDFDPQSLPIKLAIVGRPNVGKSTL
                     TNRILGEERVVVYDMPGTTRDSIYIPMERDEREYVLIDTAGVRKRGKITDAVEKFSVI
                     KTLQAIEDANVVLLVIDAREGISDQDLSLLGFILNSGRSLVIVVNKWDGLSQEVKEQV
                     KETLDFRLGFIDFARVHFISALHGSGVGNLFESVREAYDSSTRRVSTAMLTRIMTMAV
                     EDHQPPLVRGRRVKLKYAHAGGYNPPIVVIHGNQVKDLPDSYKRYLMNYFRKSLEVMG
                     TPIRIQFKEGENPYANKRNTLTPTQMRKRKRLMKHIKKSK"
     RBS             complement(11171. .11176)
                     /gene="engA"
                     /note="putative RBS for engA; RegulonDB:STMS1H002626"
     gene            complement(11283. .12471)
                     /gene="yfgL"
                     /note="synonym: STM2520"

Alignment Scores:
Pred. No.:              5.3e-132      Length:          23647
Score:                  1840.00       Matches:         367
Percent Similarity:     99.46%        Conservative:    3
Best Local Similarity:  98.66%        Mismatches:      2
Query Match:            98.61%        Indels:          0
DB:                     1             Gaps:            0

US-09-921-992-78 (1-372) x AE008814 (1-23647)
```

```
Qy          1 MetHisAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15609 ATGCATAACCAGGCTCCGATTCAACGTAGAAAATCGACACGTATTTACGTTGGGAATGTG 15550

Qy         21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrThrAsp 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15549 CCGATTGGCGATGGTGCCCCTATCGCCGTCCAGTCGATGACTAACACGCGTACCACTGAC 15490

Qy         41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15489 GTGGAAGCGACGGTTAATCAGATCAAAGCGCTGGAGCGCGTTGGCGCAGATATTGTTCGT 15430

Qy         61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsn 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15429 GTTTCTGTACCGACAATGGACGCCGCCGAAGCGTTCAAGCTTATCAAACAGCAGGTTAAC 15370

Qy         81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15369 GTCCCGCTGGTTGCCGATATTCACTTCGACTACCGCATTGCGCTGAAAGTAGCGGAATAC 15310

Qy        101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMet 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15309 GGCGTTGACTGCTTGCGTATCAACCCCGGCAATATCGGTAACGAAGAACGTATTCGTATG 15250

Qy        121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15249 GTGGTGGACTGCGCGCGCGACAAAAACATCCCTATTCGTATCGGCGTAAACGCCGGATCG 15190

Qy        141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSer 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15189 CTGGAAAAAGATCTGCAGGAAAAATACGGTGAGCCGACGCCGCAGGCGTTGCTGGAATCA 15130

Qy        161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15129 GCCATGCGCCATGTCGATCATCTCGATCGTCTCAACTTTGATCAGTTCAAAGTGAGCGTA 15070

Qy        181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15069 AAAGCGTCCGATGTATTCCTTGCCGTTGAGTCTTATCGTTTGTTGGCTAAACAGATCGAT 15010

Qy        201 GlnProLeuHisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSer 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15009 CAGCCGCTGCATCTGGGGATCACCGAAGCGGGCGGCGCGCGTAGCGGGGCGGTCAAGTCG 14950

Qy        221 AlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14949 GCGATCGGTTTAGGTCTGCTGTTATCGGAAGGTATCGGCGACACGCTGCGCGTATCGTTG 14890

Qy        241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14889 GCGGCCGACCCGGTGGAAGAGATCAAAGTCGGTTTCGATATTCTGAAGTCGCTGCGCATT 14830

Qy        261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspVal 280
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14829 CGCGCGCGCGGCATCAATTTCATCGCCTGCCCGACCTGTTCTCGCCAGGAGTTCGACGTT 14770

Qy        281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14769 ATCGGCACGGTGAACGCGCTGGAGCAGCGTCTGGAAGATATCATCACCCCGATGGACGTC 14710

Qy        301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14709 TCGATCATCGGTTGCGTAGTGAACGGGCCGGGCGAAGCGCTGGTATCGACGCTGGGCGTG 14650

Qy        321 ThrGlyGlyAsnLysLysSerGlyLeuTyrGluAspGlyValArgLysAspArgLeuAsp 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14649 ACGGGCGGCAATAAGAAAAGCGGCCTGTATGAAGACGGCGTGCGTAAGGACAGACTCGAT 14590

Qy        341 AsnAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360
              |||:::|||||||||   |||||||||:::||||||||||||||||||||||||||||||
Db      14589 AACGACGATATGATCGCGCAGCTTGAATCCCGTATTCGCGCGAAAGCAAGTCAACTTGAT 14530
```

```
Qy        361 GluAlaArgArgIleAspValGlnGlnValGluLys 372
              ||||||||||||||||||||   ||||||||||||
Db      14529 GAAGCGCGTCGGATTGACGTGCTGCAGGTTGAAAAA 14494

RESULT 14
AL627275/c
LOCUS       AL627275              145050 bp    DNA     linear   BCT 06-JUN-2002
DEFINITION  Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18,
            complete chromosome; segment 11/20.
ACCESSION   AL627275 AL513382
VERSION     AL627275.1  GI:16503698
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Typhi
  ORGANISM  Salmonella enterica subsp. enterica serovar Typhi
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 145050)
  AUTHORS   Parkhill,J., Dougan,G., James,K.D., Thomson,N.R., Pickard,D.,
            Wain,J., Churcher,C., Mungall,K.L., Bentley,S.D., Holden,M.T.G.,
            Sebaihia,M., Baker,S., Basham,D., Brooks,K., Chillingworth,T.,
            Connerton,P., Cronin,A., Davis,P., Davies,R.M., Dowd,L., White,N.,
            Farrar,J., Feltwell,T., Hamlin,N., Haque,A., Hien,T.T., Holroyd,S.,
            Jagels,K., Krogh,A., Larsen,T.S., Leather,S., Moule,S., O'Gaora,P.,
            Parry,C., Quail,M., Rutherford,K., Simmonds,M., Skelton,J.,
            Stevens,K., Whitehead,S. and Barrell,B.G.
  TITLE     Complete genome sequence of a multiple drug resistant Salmonella
            enterica serovar Typhi CT18
  JOURNAL   Nature 413 (6858), 848-852 (2001)
  MEDLINE   21534947
   PUBMED   11677608
REFERENCE   2  (bases 1 to 145050)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-OCT-2001) Submitted on behalf of the Salmonalla
            sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA, UK
COMMENT     E-mail: parkhill@sanger.ac.uk
            Notes:
            Details of S. typhi sequencing at the Sanger Centre are available
            on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/S_typhi/).
FEATURES             Location/Qualifiers
     source          1. .145050
                     /organism="Salmonella enterica subsp. enterica serovar
                     Typhi"
                     /mol_type="genomic DNA"
                     /strain="CT18"
                     /db_xref="taxon:90370"
     RBS             636. .640
                     /note="possible RBS"
     gene            647. .1003
                     /gene="STY2720"
                     /note="synonym: yffB"
     CDS             647. .1003
                     /gene="STY2720"
                     /note="Orthologue of E. coli yffB (YFFB_ECOLI); Fasta hit
                     to YFFB_ECOLI (118 aa), 81% identity in 118 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAD02682.1"
                     /db_xref="GI:16503699"
                     /db_xref="SPTREMBL:Q8Z4S3"
                     /translation="MITLYGIKNCDTIKKARRWLEEHGIDYRFHDYRVDGIDLPLLNT
                     FIAELGWQPLLNTRGTTWRKLDEGHRSGITDADSAAALMVEQPAIIKRPLLCAPGKPM
                     LLGFSESRYQQFFDEV"
     RBS             995. .1000
                     /gene="STY2720"
                     /note="possible RBS"
     gene            1007. .2134
                     /gene="STY2721"
                     /note="synonym: dapE"
     CDS             1007. .2134
                     /gene="STY2721"
                     /note="Orthologue of E. coli dapE (DAPE_ECOLI); Fasta hit
                     to DAPE_ECOLI (375 aa), 94% identity in 375 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="succinyl-diaminopimelate desuccinylase"
                     /protein_id="CAD02683.1"
                     /db_xref="GI:16503700"
                     /db_xref="SPTREMBL:Q8Z4S2"
                     /translation="MSCPVIELTQQLIRRPSLSPDDVGCQALMIERLRKIGFTIEHMD
                     FGDTQNFWAWRGRGETLAFAGHTDVVPAGDVDRWINPPFEPTIRDGMLFGRGAADMKG
                     SLAAMVVAAERFVAQHPHHRGRLAFLITSDEEASAKNGTVKVVEALMARNERLDYCLV
                     GEPSSTEIVGDVVKNGRRGSLTCNLTIHGVQGHVAYPHLADNPVHRAAPFLNELVAIE
                     WDRGNDFFPATSMQIANIQAGTGSNNVIPGELFVQFNFRFSTELTDEMIKERVHALLE
                     KHQLRYTVDWWLSGQPFLTARGKLVDAVVNAIEHYNEIKPQLLTTGGTSDGRFIARMG
                     AQVVELGPVNATIHKINECVNAADLQLLARMYQRIMEQLVA"
     misc_feature    1031. .1978
                     /gene="STY2721"
                     /note="Pfam match to entry PF01546 Peptidase_M20,
                     Peptidase family M20/M25/M40, score 315.30, E-value
                     7.4e-91"
     misc_feature    1187. .1216
                     /gene="STY2721"
                     /note="PS00758 ArgE / dapE / ACY1 / CPG2 / yscS family
                     signature 1"
     misc_feature    1295. .1414
                     /gene="STY2721"
                     /note="PS00759 ArgE / dapE / ACY1 / CPG2 / yscS family
                     signature 2"
     RBS             2150. .2153
                     /note="possible RBS"
     gene            2162. .2362
                     /gene="STY2722"
     CDS             2162. .2362
                     /gene="STY2722"
                     /note="Similar to Escherichia coli similar to TR:P77259
                     (EMBL:D90875) (66 aa) fasta scores: E(): 1e-27, 95.5% id
                     in 66 aa. Contains hydrophobic, possible membrane-spanning
                     region"
                     /codon_start=1
                     /transl_table=11
                     /product="putative membrane protein"
                     /protein_id="CAD02684.1"
                     /db_xref="GI:16503701"
                     /db_xref="SPTREMBL:Q8XF02"
                     /translation="MDWLAKYWWILVLVFLVGVLLNVIKDLKRIDHKKFLANKPELPP
                     HRDFNDKWDDEDDWPKKDQPKK"
     gene            complement(2389. .4404)
                     /gene="STY2723"
     CDS             complement(2389. .4404)
                     /gene="STY2723"
                     /note="Orthologue of E. coli YPFI_ECOLI; Fasta hit to
                     YPFI_ECOLI (671 aa), 72% identity in 662 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAD02685.1"
                     /db_xref="GI:16503702"
                     /db_xref="SPTREMBL:Q8Z4S1"
                     /translation="MSDIDALRALTSQMTQEGIRRLLVISGDAAWCRERAEAIRAVLP
                     GDWWVAPDAPAQPRCTPQALQTLLGREFRHAIFDAWQGFDAAAFAALSGTLQAGSWLL
                     LLMPPYETWESRPDTDSLRWSDCAQPIPTPQFAQHLKRTLSRDPQTLLWRQRQPFCWP
                     SYPSRERWRPATGEPQPEQAAILSRLREMPPGVATVIAPRGRGKSALAGQFISQMAGT
                     AIVTAPAKTATDILAAFAGERFCFMAPDALLASGARADWLVVDEAAAIPAPLLLQLVS
                     RFPRILLTTTVQGYEGTGRGFLLKFCARFPQLHRFTLRQPVRWAPECPLENIVSEALI
                     FDDEAFAQAPHGGIAISAFYQQAWGKTPALPRAVYQLLSGAHYRTSPLDLRRMMDAPG
                     QHFLQATANNRVAGALWLVEEGGLSAELSQAVWCGFRRRPRGNLVAQSLAAHGSDPLAA
                     TLVGRRVSRIAVHPARQREGIGQQLIACACEQAAQCDYLSVSFGYTPELWRFWQRCGF
                     VLVRMGSHREASSGCYTAMALLPLSDAGQRLAQQEHRRLRRDADILTQWNGEAIPLAA
                     LREQALNDEDWRELVGFAFAHRPLLTSLGCLHRLLQYSALPLPALRGRLEEKASDAEL
                     CARLRISGRKALLALQRAQAAQALIALDAGRTQRLRDVMPGGGDHAG"
     misc_feature    complement(2878. .3207)
                     /gene="STY2723"
```

```
                     /note="Pfam match to entry PF00583 Acetyltransf,
                     Acetyltransferase (GNAT) family, score 32.50, E-value
                     9.7e-06"
     misc_feature    complement(3790. .3813)
                     /gene="STY2723"
                     /note="PS00017 ATP/GTP-binding site motif A (P-loop)"
     gene            complement(4420. .5283)
                     /gene="STY2724"
     CDS             complement(4420. .5283)
                     /gene="STY2724"
                     /note="Orthologue of E. coli P76563; Fasta hit to P76563
                     (287 aa), 92% identity in 286 aa overlap. Contains
                     hydrophobic, possible membrane-spanning region"
                     /codon_start=1
                     /transl_table=11
                     /product="putative membrane protein"
                     /protein_id="CAD02686.1"
                     /db_xref="GI:16503703"
                     /db_xref="SPTREMBL:Q8Z4S0"
                     /translation="MRWQGRRESNNVEDRRNRPGGPSLGGPGFRLPRGKGGIILLVVV
                     VVAGYYGVDLTGLLTGQPVSQQQSTRSISPNDDEAAKFTSVILATTEDTWGQLFQKMG
                     RGYQQPKLVMYRGMTRTGCGAGQSVMGPFYCPADGTVYIDLSFYDVMKNKLGADGDFA
                     QGYVIAHEVGHHVQKLLGIEPKVRQLQQNASQTEVNRLSVRMELQADCFAGVWGHSMQ
                     QQGVLEAGDLEEALNAAQAIGDDRLQQQGQGRVVPDSFTHGTSEQRYSWFKRGFDSGD
                     PAQCNTFGKNF"
     RBS             complement(5287. .5292)
                     /note="possible RBS"
     gene            complement(5415. .6128)
                     /gene="STY2725"
                     /note="synonym: purC"
     CDS             complement(5415. .6128)
                     /gene="STY2725"
                     /note="Orthologue of E. coli purC (PUR7_ECOLI); Fasta hit
                     to PUR7_ECOLI (237 aa), 95% identity in 237 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="phosphoribosylaminoimidazole-succinocarboxamide
                     synthase"
                     /protein_id="CAD02687.1"
                     /db_xref="GI:16503704"
                     /db_xref="SPTREMBL:Q8XF69"
                     /translation="MQKQAELYRGKAKTVYSTENPDLLVLEFRNDTSAGDGARIEQFD
                     RKGMVNNKFNHFIMTKLAEAGIPTQMERLLSDTECLVKKLEMVPVECVVRNRAAGSLV
                     KRLGVEEGMELNPPIFDLFLKNDALHDPMVNSSYCETFGWVSQENLARMKELTYKAND
                     VLKKLFDDAGLILVDFKLEFGLYKGEVVLGDEFSPDGSRLWDKETLDKMDKDRFRQSL
                     GGLIEAYEAVAHRLGVKLD"
     misc_feature    complement(5421. .6122)
                     /gene="STY2725"
                     /note="Pfam match to entry PF01259 SAICAR_synt, SAICAR
                     synthetase, score 414.20, E-value 1.2e-120"
     misc_feature    complement(5586. .5612)
                     /gene="STY2725"
                     /note="PS01058 SAICAR synthetase signature 2"
     misc_feature    complement(5829. .5873)
                     /gene="STY2725"
                     /note="PS01057 SAICAR synthetase signature 1"
     RBS             complement(6137. .6141)
                     /note="possible RBS"
     gene            complement(6303. .7337)
                     /gene="STY2726"

Alignment Scores:
Pred. No.:              2.9e-130        Length:         145050
Score:                  1831.00         Matches:        365
Percent Similarity:     99.19%          Conservative:   4
Best Local Similarity:  98.12%          Mismatches:     3
Query Match:            98.12%          Indels:         0
DB:                     1               Gaps:           0

US-09-921-992-78 (1-372) x AL627275 (1-145050)

Qy          1 MetHisAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db      64624 ATGCATAACCAGGCTCCGATTCAACGTAGAAAATCGACACGTATTTACGTTGGGAATGTG 64565

Qy         21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrThrAsp 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db      64564 CCGATTGGCGATGGCGCCCCTATCGCCGTCCAGTCGATGACTAACACGCGTACCACTGAC 64505

Qy         41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db      64504 GTGGAAGCGACGGTTAATCAGATCAAAGCGCTGGAGCGCGTTGGCGCAGATATTGTTCGT 64445

Qy         61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsn 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::

Db      64444 GTTTCTGTACCGACCATGGACGCCGCCGAAGCGTTCAAGCTTATCAAACAGCAGGTTAGC 64385

Qy         81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db      64384 GTCCCGCTGGTTGCCGATATTCACTTCGACTACCGCATTGCGCTGAAAGTAGCGGAATAC 64325

Qy        101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMet 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db      64324 GGCGTTGACTGCTTGCGTATTAACCCCGGCAATATCGGTAACGAAGAACGCATTCGCATG 64265

Qy        121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db      64264 GTGGTGGACTGCGCGCGCGACAAAAACATCCCTATCCGTATCGGCGTAAATGCCGGATCG 64205

Qy        141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuGluSer 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db      64204 CTGGAAAAAGATCTGCAGGAAAAATACGGAGAGCCGACGCCGCAGGCGCTGCTGGAATCA 64145

Qy        161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db      64144 GCCATGCGCCATGTCGATCATCTCGATCGTCTCAACTTTGATCAGTTCAAAGTGAGCGTA 64085

Qy        181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db      64084 AAAGCGTCCGATGTATTCCTTGCCGTTGAGTCTTATCGTTTGTTGGCTAAACAGATCGAT 64025

Qy        201 GlnProLeuHisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSer 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db      64024 CAGCCGCTGCATCTGGGGATCACCGAAGCGGGCGGCGCGCGTAGCGGGGCGGTTAAGTCG 63965

Qy        221 AlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db      63964 GCGATCGGTTTAGGTCTGCTGTTATCGGAAGGCATCGGCGACACGCTGCGCGTATCGTTG 63905

Qy        241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db      63904 GCGGCCGACCCGGTGGAAGAGATCAAAGTCGGTTTCGATATCCTGAAGTCGCTGCGTATT 63845

Qy        261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspVal 280
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db      63844 CGCGCGCGCGGCATCAATTTCATCGCCTGCCCGACCTGTTCTCGTCAGGAGTTCGACGTT 63785

Qy        281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db      63784 ATCGGCACGGTGAACGCGCTGGAGCAGCGTCTGGAAGATATCATCACCCCGATGGACGTC 63725

Qy        301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db      63724 TCGATCATCGGTTGTGTGGTGAACGGGCCGGGCGAAGCGCTGGTATCGACGCTGGGCGTG 63665

Qy        321 ThrGlyGlyAsnLysLysSerGlyLeuTyrGluAspGlyValArgLysAspArgLeuAsp 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db      63664 ACGGGCGGCAATAAGAAAAGCGGCCTGTATGAAGACGGCGTGCGTAAGGACAGACTCGAT 63605

Qy        341 AsnAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360
              |||:::|||||||||   ||||||||||:::|||||||||||||||   ||||||||||||

Db      63604 AACGACGATATGATCGCGCAGCTTGAATCCCGTATTCGCGCGAAAGTAAGTCAACTTGAT 63545

Qy        361 GluAlaArgArgIleAspValGlnGlnValGluLys 372
              ||||||||||||||||||||   ||||||||||||

Db      63544 GAAGCGCGTCGGATTGACGTGCTGCAGGTTGAAAAA 63509
```

```
RESULT 15
AE016835
LOCUS       AE016835              300431 bp    DNA     linear   BCT 21-MAR-2003
DEFINITION  Salmonella enterica subsp. enterica serovar Typhi Ty2, section 2 of
            16 of the complete genome.
ACCESSION   AE016835 AE014613
VERSION     AE016835.1  GI:29136420
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Typhi Ty2
  ORGANISM  Salmonella enterica subsp. enterica serovar Typhi Ty2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 300431)
  AUTHORS   Deng,W., Liou,S.R., Plunkett III,G., Mayhew,G.F., Rose,D.J.,
            Burland,V., Kodoyianni,V., Schwartz,D.C. and Blattner,F.R.
  TITLE     Comparative Genomics of Salmonella enterica Serovar Typhi Strains
            Ty2 and CT18
  JOURNAL   J. Bacteriol. 185 (7), 2330-2337 (2003)
  MEDLINE   22531367
   PUBMED   12644504
REFERENCE   2  (bases 1 to 300431)
  AUTHORS   Deng,W., Liou,S.-R., Plunkett,G. III, Mayhew,G.F., Rose,D.J.,
            Burland,V., Kodoyianni,V., Schwartz,D.C. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-SEP-2002) Laboratory of Genetics, University of
            Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .300431
                     /organism="Salmonella enterica subsp. enterica serovar
                     Typhi Ty2"
                     /mol_type="genomic DNA"
                     /strain="Ty2"
                     /serovar="Typhi"
                     /sub_species="enterica"
                     /db_xref="taxon:209261"
     gene            complement(67. .486)
                     /gene="yfiG"
                     /locus_tag="t0261"
     CDS             complement(67. .486)
                     /gene="yfiG"
                     /locus_tag="t0261"
                     /note="corresponds to STY2842 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="thioredoxin 2"
                     /protein_id="AAO67987.1"
                     /db_xref="GI:29136421"
                     /translation="MNTVCTHCQAINRIPGDRLQDAAKCGRCGHELFDGEVINATGET
                     LDKLLKDDLPVVIDFWAPWCGPCRNFAPIFEDVAEERSGKVRFVKVNTEAERELSARF
                     GIRSIPMIMIFKHGQVVDMLNGAVPKAPFDSWLNEAL"
     gene            690. .1727
                     /gene="yfiF"
                     /locus_tag="t0262"
     CDS             690. .1727
                     /gene="yfiF"
                     /locus_tag="t0262"
                     /note="corresponds to STY2841 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="putative RNA methyltransferase"
                     /protein_id="AAO67988.1"
                     /db_xref="GI:29136422"
                     /translation="MNDELKNKSGKVKVMYVRSDDDSDKRTHNPRTGKGGGRPAKSRT
                     DGGRRPARDERNNQSRDRKHETSPWRTVSRAPGDETPEKVDHGGISGKSFIDPEVLRR
                     QRAEETRVYGENACQALFQSRPDAIVRAWFIQSVTPRFKEALRWMAANRKAYHVVDEA
                     ELAKASGTEHHGGVCFLIKKRNGTTVKQWVKQAADQDCVLALEDVANPHNLGGMMRSC
                     AHFGVKGVVVQDAALLESGAAIRTAEGGAEHVQPITGESIVDVLDDFRQAGYTVVTTS
                     SDRGQALFSTTLPEKMVLVLGREYDYLPEAAREPDDLCVKINGTGNVESLNVSVATGV
                     LLAEWWRQNKA"
     gene            complement(1843. .2532)
                     /gene="ung"
                     /locus_tag="t0263"
     CDS             complement(1843. .2532)
                     /gene="ung"
                     /locus_tag="t0263"
                     /note="corresponds to STY2840 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="uracil-DNA glycosylase"
                     /protein_id="AAO67989.1"
                     /db_xref="GI:29136423"
                     /translation="MATELTWHDVLADEKQQPYFINTLHTVAGERQSGITVYPPQKDV
                     FNAFRFTELGDVKVVILGQDPYHGPGQAHGLAFSVRPGIAPPPSLVNMYKELEASIPG
                     FVRPAHGYLESWARQGVLLLNTVLTVRAGQAHSHASLGWETFTDKVISLINQHREGVV
                     FLLWGSHAQKKGAIIDPQRHHILKAPHPSPLSAHRGFFGCNHFALTNQWLEQHGEKTI
                     DWTPVLPAESE"
     gene            2851. .3234
                     /gene="yfiD"
                     /locus_tag="t0264"
     CDS             2851. .3234
                     /gene="yfiD"
                     /locus_tag="t0264"
                     /note="corresponds to STY2839 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAO67990.1"
                     /db_xref="GI:29136424"
                     /translation="MITGIQITKAANDDLLNSFWLLDSEKGEARCIVAKSGFAEDEVV
                     AVSKLGEIEYREIPMEVKPEVRVEGGQHLNVNVLRRETLEDAVKHPEKYPQLTIRVSG
                     YAVRFNSLTPEQQRDVIARTFTESL"
     gene            complement(3296. .3883)
                     /gene="yfiK"
                     /locus_tag="t0265"
     CDS             complement(3296. .3883)
                     /gene="yfiK"
                     /locus_tag="t0265"
                     /note="corresponds to STY2838 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="putative membrane protein"
                     /protein_id="AAO67991.1"
                     /db_xref="GI:29136425"
                     /translation="MTPMLLSAFWTYTLITALTPGPNNILALSAATAHGFRQSIRVLA
                     GMSLGFLVVMLLCAGIAFSLAVIDPAIIHLLSWVGAAYILWLAWKIATSPAADEKVRP
                     KPVGFWVSFGLQFVNVKIILYGITALSTFVLPQTQALNWVIGVSILLALIGTFGNVCW
                     ALAGHLFQRAFRHYGRQLNIILALLLVYCAVRIFY"
     gene            3986. .4868
                     /gene="yfiE"
                     /locus_tag="t0266"
                     /note="corresponds to STY2837 from Accession AL513382:
                     Salmonella typhi CT18"
                     /pseudo
     gene            complement(4901. .6235)
                     /gene="srmB"
                     /locus_tag="t0267"
     CDS             complement(4901. .6235)
                     /gene="srmB"
                     /locus_tag="t0267"
                     /note="corresponds to STY2836 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="ATP-dependent RNA helicase SrmB"
                     /protein_id="AAO67992.1"
                     /db_xref="GI:29136426"
                     /translation="MTVTTFSELELDESLLDALQDKGFTRPTAIQAAAIPPALDGRDV
                     LGSAPTGTGKTAAYLLPALQHLLDFPRKKSGPPRILILTPTRELAMQVADHARELAKH
                     THLDIATITGGVAYMNHAEVFSENQDIVVATTGRLLQYIKEENFDCRAVETLILDEAD
                     RMLDMGFAQDIEHIAGETRWRKQTMLFSATLEGDAIKDFAERLLEDPVEVSANPSTRE
```

```
                RKKIHQWYYRADNFEHKVALLKHLLKQDDATRSIVFVRKRERVHELAETLRLAGINNC
                YLEGEMAQIKRNEGIKRLTDGRVNVLVATDVAARGIDIPDVSHVINFDMPRSGDTYLH
                RIGRTGRAGRKGTAISLVEAHDHLLLLKIGRYIEEPLKARVIDELRPTTRAPSEKLTG
                KPSKKVLAKRAEKKKEKEKEKPRVKKRYRDTKNIGKRRKPSGTKMQEQSSEE"
    gene        6365. .7102
                /gene="yfiC"
                /locus_tag="t0268"
    CDS         6365. .7102
                /gene="yfiC"
                /locus_tag="t0268"
                /note="corresponds to STY2835 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="AAO67993.1"
                /db_xref="GI:29136427"
                /translation="MSQSGSVLRRNGFTFKQFFVAHDRCAMKVGTDGILLGAWAPVAD
                VKRILDIGTGSGLLALMLAQRTDDSVPVDAVELDAGAAMQAQENVAHSPWPHRITVHT
                DDIQRWAPRQTVRFDLIISNPPYYEPGVECSTPQREQARYTATLDHQTLLAIAADCIT
                EDGFFCVVLPEQIGNAFTQQALNMGWHLRLRTDVAENEARLPHRVLLAFSPQAGECFS
                DRLVIRGSDQNYSESYTALTQAFYLFM"
    gene        complement(7087. .8709)
                /gene="nadB"
                /locus_tag="t0269"
    CDS         complement(7087. .8709)
                /gene="nadB"
                /locus_tag="t0269"
                /note="quinolinate synthetase B; corresponds to STY2834
                from Accession AL513382: Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="L-aspartate oxidase"
                /protein_id="AAO67994.1"
                /db_xref="GI:29136428"
                /translation="MMTTPELSCDVLIIGSGAAGLSLALRLAEKHKVIVLSKGPVSEG
                STFYAQGGIAAVFDETDSIASHVEDTLIAGAGICDRHAVEFVASNARTCVQWLIDQGV
                LFDTHVQPNGKESYHLTREGGHSHRRILHAADATGKEVETTLVSRAQNHPNIQVLERS
                NAVDLIISDKIGLPGPRRVVGAWIWNRNKEWVETCHAKSVVLATGGASKVYQYTTNPD
                ISSGDGIAMAWRAGCRVANLEFNQFHPTALYHPQARNFLLTEALRGEGAYLKRPDGSR
                FMPDVDERGELAPRDIVARAIDHEMKRLGADCMFLDISHKPDDFVRQHFPMIYAKLLD
                LGMDLTKEPIPIVPAAHYTCGGVVVDDYGRTDVDGLYAIGEVSYTGLHGANRMASNSL
                LECLVYGWSAAMDIDRRMPSVHSVDVLPAWDESRVENADERVVIQHNWHELRLLMWDY
                VGIVRTTKRLERALRRITMLQQEIDEYYANFRVSNNLLELRNLVQVAELIVRCAMMRK
                ESRGLHFTLDYPQQLAESGPSILSPLTPHINR"
    gene        9134. .9709
                /gene="rpoE"
                /locus_tag="t0270"
    CDS         9134. .9709
                /gene="rpoE"
                /locus_tag="t0270"
                /note="sigma-24; corresponds to STY2833 from Accession
                AL513382: Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="RNA polymerase sigma-E factor"
                /protein_id="AAO67995.1"

Alignment Scores:
Pred. No.:                7.58e-130       Length:         300431
Score:                    1831.00         Matches:        365
Percent Similarity:       99.19%          Conservative:   4
Best Local Similarity:    98.12%          Mismatches:     3
Query Match:              98.12%          Indels:         0
DB:                       1               Gaps:           0

US-09-921-992-78 (1-372) x AE016835 (1-300431)

Qy          1 MetHisAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83116 ATGCATAACCAGGCTCCGATTCAACGTAGAAAATCGACACGTATTTACGTTGGGAATGTG 83175

Qy         21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrThrAsp 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83176 CCGATTGGCGATGGCGCCCCTATCGCCGTCCAGTCGATGACTAACACGCGTACCACTGAC 83235

Qy         41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83236 GTGGAAGCGACGGTTAATCAGATCAAAGCGCTGGAGCGCGTTGGCGCAGATATTGTTCGT 83295

Qy         61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsn 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db      83296 GTTTCTGTACCGACCATGGACGCCGCCGAAGCGTTCAAGCTTATCAAACAGCAGGTTAGC 83355

Qy         81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83356 GTCCCGCTGGTTGCCGATATTCACTTCGACTACCGCATTGCGCTGAAAGTAGCGGAATAC 83415

Qy        101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMet 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83416 GGCGTTGACTGCTTGCGTATTAACCCCGGCAATATCGGTAACGAAGAACGCATTCGCATG 83475

Qy        121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83476 GTGGTGGACTGCGCGCGCGACAAAAACATCCCTATCCGTATCGGCGTAAATGCCGGATCG 83535

Qy        141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSer 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83536 CTGGAAAAAGATCTGCAGGAAAAATACGGAGAGCCGACGCCGCAGGCGCTGCTGGAATCA 83595

Qy        161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83596 GCCATGCGCCATGTCGATCATCTCGATCGTCTCAACTTTGATCAGTTCAAAGTGAGCGTA 83655

Qy        181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83656 AAAGCGTCCGATGTATTCCTTGCCGTTGAGTCTTATCGTTTGTTGGCTAAACAGATCGAT 83715

Qy        201 GlnProLeuHisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSer 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83716 CAGCCGCTGCATCTGGGGATCACCGAAGCGGGCGGCGCGCGTAGCGGGGCGGTTAAGTCG 83775

Qy        221 AlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83776 GCGATCGGTTTAGGTCTGCTGTTATCGGAAGGCATCGGCGACACGCTGCGCGTATCGTTG 83835

Qy        241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83836 GCGGCCGACCCGGTGGAAGAGATCAAAGTCGGTTTCGATATCCTGAAGTCGCTGCGTATT 83895

Qy        261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspVal 280
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83896 CGCGCGCGCGGCATCAATTTCATCGCCTGCCCGACCTGTTCTCGTCAGGAGTTCGACGTT 83955

Qy        281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83956 ATCGGCACGGTGAACGCGCTGGAGCAGCGTCTGGAAGATATCATCACCCCGATGGACGTC 84015

Qy        301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      84016 TCGATCATCGGTTGTGTGGTGAACGGGCCGGGCGAAGCGCTGGTATCGACGCTGGGCGTG 84075

Qy        321 ThrGlyGlyAsnLysLysSerGlyLeuTyrGluAspGlyValArgLysAspArgLeuAsp 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      84076 ACGGGCGGCAATAAGAAAAGCGGCCTGTATGAAGACGGCGTGCGTAAGGACAGACTCGAT 84135

Qy        341 AsnAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360
              |||:::|||||||||   ||||||||||:::|||||||||||||||   ||||||||||||
Db      84136 AACGACGATATGATCGCGCAGCTTGAATCCCGTATTCGCGCGAAAGTAAGTCAACTTGAT 84195

Qy        361 GluAlaArgArgIleAspValGlnGlnValGluLys 372
              ||||||||||||||||||||   ||||||||||||
Db      84196 GAAGCGCGTCGGATTGACGTGCTGCAGGTTGAAAAA 84231

Search completed: August 15, 2003, 00:56:40
```

Job time : 5342 secs

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 14, 2003, 21:49:35 ; Search time 356 Seconds
(without alignments)
2820.761 Million cell updates/sec

Title: US-09-921-992-78
Perfect score: 1866
Sequence: 1 MHNQAPIQRRKSTRIYVGNV.....RAKASQLDEARRIDVQQVEK 372

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=x1h
-Q=/cgn2_1/USPTO_spool/US09921992/runat_12082003_094910_17323/app_query.fasta_1.519
-DB=N_Geneseq_19Jun03 -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09921992_@CGN_1_1_312_@runat_12082003_094910_17323 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1866	100.0	1119	21	AAA95478	E. coli essential
	2	1866	100.0	1119	21	AAA88705	E. coli FUN essent
	3	1866	100.0	1119	22	AAC82653	E. coli gcpE DNA.
	4	1866	100.0	1119	24	AAD31203	Escherichia coli g
	5	1603.5	85.9	1830121	17	AAT42063	Haemophilus influe
	6	1282	68.7	640681	24	ABA92787	Buchnera sp. genom
	7	1166	62.5	33140	22	AAF28536	Genomic fragment #
	8	868	46.5	1107	24	ABQ67966	Listeria monocytog
	9	868	46.5	1107	24	ABQ69993	Listeria monocytog
	10	868	46.5	2944528	24	ABA03041	Listeria monocytog
	11	861.5	46.2	6157	24	ABQ70939	Listeria monocytog
	12	822	44.1	1083	24	ABK74401	Bacillus lichenifo
	13	814.5	43.6	349980	24	ABQ81842	Bifidobacterium lo
	14	812	43.5	994	24	ABQ68324	Listeria monocytog
	15	794	42.6	1180	19	AAX14017	H. pylori GHPO 76
c	16	785.5	42.1	35829	23	AAS59573	Propionibacterium
	17	761	40.8	1134	22	AAH67170	C glutamicum codin
	18	761	40.8	1179	25	ACA00649	C. glutamicum deri
c	19	761	40.8	349980	22	AAH68530	C glutamicum codin
c	20	761	40.8	349980	22	AAH68531	C glutamicum codin
	21	751.5	40.3	1161	25	ABZ71126	Mycobacterium tube
	22	751.5	40.3	1164	22	AAH52056	Mycobacterium tube
c	23	751.5	40.3	4403765	22	AAI99683	Mycobacterium tube
c	24	751.5	40.3	4411529	22	AAI99682	Mycobacterium tube
	25	682	36.5	5484	23	AAS73006	DNA encoding novel
	26	534	28.6	1263	25	ABZ39756	N. gonorrhoeae nuc
c	27	527	28.2	60873	21	AAA81469	N. meningitidis pa
	28	527	28.2	349980	21	AAF21610	Neisseria meningit
	29	527	28.2	1437668	21	AAA81490	N. meningitidis B
	30	510	27.3	1857	23	AAS90066	DNA encoding novel
	31	490	26.3	1038602	20	AAZ01425	Complete genome se
c	32	470.5	25.2	273254	21	AAC81914	Chlamydia pneumoni
	33	470.5	25.2	1230025	20	AAX91990	Nucleotide sequenc
	34	462	24.8	2520	24	AAD31201	Arabidopsis thalia
	35	460	24.7	2109	22	AAC82654	P. falciparum gcpE
	36	446	23.9	3147	21	AAX82996	P. falciparum gcpE
	37	428.5	23.0	507	21	AAC75580	Human ORFX ORF1135
	38	428.5	23.0	507	24	ABN76496	Human ORF1443 cDNA
	39	383	20.5	435	24	ABX66784	Helicobacter pylor
	40	339	18.2	670	24	AAD31221	Zea mays partial g
	41	326	17.5	596	24	AAD31222	Zea mays partial g
c	42	321	17.2	584	24	AAD31220	Zea mays partial g
c	43	300	16.1	4467	10	AAN92428	Sequence encoding
	44	292	15.6	33675	24	AAD31202	Oryza sativa gcpE
c	45	277.5	14.9	2535	20	AAX20674	Polynucleotide seq

ALIGNMENTS

RESULT 1
AAA95478
ID AAA95478 standard; DNA; 1119 BP.
XX
AC AAA95478;
XX
DT 27-FEB-2001 (first entry)
XX
DE E. coli essential gene gcpE.
XX
KW Bacterial growth; inhibitor; ygbB; yfhC; yacE; ychB; yejD; yrfI;
KW yggJ; yjeE; yiaO; yrdC; yhbC; ygbP; ybeY; gcpE; kdtB; pfs; ycaJ;
KW b1808; yeaA; yagF; b1983; yidD; yceG; yjbC; antibacterial compound;
KW H. influenza; S. pneumoniae; H. pylori; B. burgdorferi; pesticide;
KW M. tuberculosis; antibiotic; ds.
XX
OS Escherichia coli.

XX
PN WO200061793-A2.
XX
PD 19-OCT-2000.
XX
PF 07-APR-2000; 2000WO-EP03135.
XX
PR 09-APR-1999; 99EP-0107031.
PR 04-FEB-2000; 2000EP-0102111.
XX
PA (GPCB-) GPC BIOTECH AG.
XX
PI Loferer H, Jacobi A;
XX
DR WPI; 2000-687048/67.
XX
PT Identifying antibacterial compounds, comprises identifying an
PT antagonist or inhibitor of the expression of a gene encoding a
PT polypeptide essential for bacterial growth or survival -
XX
PS Claim 1; Fig 1; 75pp; English.
XX
CC The present invention relates to antagonists and inhibitors of 24
CC bacterial genes and proteins. The proteins are thought to be essential
CC for growth in several species of bacteria (including S. pneumoniae, B.
CC burgdorferi, H. influenza and H. pylori). The proteins and coding
CC sequences shown in the specification can be used to identify antagonists
CC and inhibitors which can be used in disease treatment and pesticides. In
CC particular, they can be used against M. tuberculosis. The present
CC sequence is one of the genes of the invention.
XX
SQ Sequence 1119 BP; 272 A; 279 C; 317 G; 251 T; 0 other;

Alignment Scores:

Pred. No.:	7.12e-184	Length:	1119
Score:	1866.00	Matches:	372
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	21	Gaps:	0

US-09-921-992-78 (1-372) x AAA95478 (1-1119)

```
Qy          1 MetHisAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnVal  20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCATAACCAGGCTCCAATTCAACGTAGAAAATCAACACGTATTTACGTTGGGAATGTG  60

Qy         21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrThrAsp  40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCGATTGGCGATGGTGCTCCCATCGCCGTACAGTCCATGACCAATACGCGTACGACAGAC 120

Qy         41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg  60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GTCGAAGCAACGGTCAATCAAATCAAGGCGCTGGAACGCGTTGGCGCTGATATCGTCCGT 180

Qy         61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsn  80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GTATCCGTACCGACGATGGACGCGGCAGAAGCGTTCAAACTCATCAAACAGCAGGTTAAC 240

Qy         81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTGCCGCTGGTGGCTGACATCCACTTCGACTATCGCATTGCGCTGAAAGTAGCGGAATAC 300

Qy        101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMet 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGCGTCGATTGTCTGCGTATTAACCCTGGCAATATCGGTAATGAAGAGCGTATTCGCATG 360

Qy        121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GTGGTTGACTGTGCGCGCGATAAAAACATTCCGATCCGTATTGGCGTTAACGCCGGATCG 420

Qy        141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSer 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTGGAAAAAGATCTGCAAGAAAAGTATGGCGAACCGACGCCGCAGGCGTTGCTGGAATCT 480

Qy        161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GCCATGCGTCATGTTGATCATCTCGATCGCCTGAACTTCGATCAGTTCAAAGTCAGCGTG 540

Qy        181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AAAGCGTCTGACGTCTTCCTCGCTGTTGAGTCTTATCGTTTGCTGGCAAAACAGATCGAT 600

Qy        201 GlnProLeuHisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSer 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CAGCCGTTGCATCTGGGGATCACCGAAGCCGGTGGTGCGCGCAGCGGGGCAGTAAAATCC 660

Qy        221 AlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GCCATTGGTTTAGGTCTGCTGCTGTCTGAAGGCATCGGCGACACGCTGCGCGTATCGCTG 720

Qy        241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GCGGCCGATCCGGTCGAAGAGATCAAAGTCGGTTTCGATATTTTGAAATCGCTGCGTATC 780

Qy        261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspVal 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CGTTCGCGAGGGATCAACTTCATCGCCTGCCCGACCTGTTCGCGTCAGGAATTTGATGTT 840

Qy        281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 ATCGGTACGGTTAACGCGCTGGAGCAACGCCTGGAAGATATCATCACTCCGATGGACGTT 900

Qy        301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TCGATTATCGGCTGCGTGGTGAATGGCCCAGGTGAGGCGCTGGTTTCTACACTCGGCGTC 960

Qy        321 ThrGlyGlyAsnLysLysSerGlyLeuTyrGluAspGlyValArgLysAspArgLeuAsp 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ACCGGCGGCAACAAGAAAAGCGGCCTCTATGAAGATGGCGTGCGCAAAGACCGTCTGGAC 1020

Qy        341 AsnAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AACAACGATATGATCGACCAGCTGGAAGCACGCATTCGTGCGAAAGCCAGTCAGCTGGAC 1080

Qy        361 GluAlaArgArgIleAspValGlnGlnValGluLys 372
              ||||||||||||||||||||||||||||||||||||
Db       1081 GAAGCGCGTCGAATTGACGTTCAGCAGGTTGAAAAA 1116
```

RESULT 2
AAA88705
ID AAA88705 standard; DNA; 1119 BP.
XX
AC AAA88705;
XX
DT 05-FEB-2001 (first entry)
XX
DE E. coli FUN essential gene gcpE.
XX
KW FUN gene; gcpE gene; essential gene; antibacterial; antibiotic;
KW screening; infection; therapy; antagonist; surrogate marker; ds.
XX
OS Escherichia coli.
XX
PN EP1043403-A1.
XX
PD 11-OCT-2000.
XX
PF 09-APR-1999; 99EP-0107031.
XX
PR 09-APR-1999; 99EP-0107031.
XX
PA (GPCG-) GPC GENOME PHARM CORP AG.
XX
PI Not given;
XX

```
DR   WPI; 2000-640125/62.
XX
PT   Identifying antagonists of the expression of gene encoding bacterial
PT   growth polypeptide useful for treating bacterial infections or
PT   diseases, by evaluating transcription of the gene in the presence of
PT   test molecule -
XX
PS   Claim 1; Page 25; 55pp; English.
XX
CC   The present sequence is that of the Escherichia coli FUN gene
CC   gcpE, which encodes a protein that is essential for bacterial
CC   growth or survival. gcpE is 1 of 22 E. coli genes (see
CC   AAA88692-713) identified as being essential (there is no deletion
CC   genotype). These 22 genes fulfil criteria for being attractive
CC   antibacterial targets: hypothetical open reading frames coding for
CC   essential functions (mutation is lethal for growth in rich media);
CC   broad conservation (orthologues are present in a wide range of
CC   bacteria, including Haemophilus influenzae, Streptococcus pneumoniae,
CC   Helicobacter pylori and Borrelia burgdorferi); and low toxicity
CC   potential in higher organisms (mostly no orthologues were identified
CC   in Saccharomyces cerevisiae). An antagonist or inhibitor of the
CC   expression of an essential gene or of its function provides the key
CC   for antibacterial therapy. The invention provides methods for
CC   identifying such antagonists or inhibitors. These involve
CC   contacting a bacterial cell comprising an essential gene with a
CC   candidate antagonist or inhibitor, and testing whether contact leads
CC   to cell growth inhibition and/or cell death. The method allows the
CC   development of new broad spectrum antibiotics. A conditional mutant
CC   of an essential gene can be used to induce a lethal phenotype in
CC   bacteria for the analysis of surrogate markers.
XX
SQ   Sequence 1119 BP; 272 A; 279 C; 317 G; 251 T; 0 other;

Alignment Scores:
Pred. No.:                7.12e-184      Length:          1119
Score:                    1866.00        Matches:         372
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       21             Gaps:            0

US-09-921-992-78 (1-372) x AAA88705 (1-1119)

Qy          1 MetHisAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCATAACCAGGCTCCAATTCAACGTAGAAAATCAACACGTATTTACGTTGGGAATGTG 60

Qy         21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrThrAsp 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCGATTGGCGATGGTGCTCCCATCGCCGTACAGTCCATGACCAATACGCGTACGACAGAC 120

Qy         41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GTCGAAGCAACGGTCAATCAAATCAAGGCGCTGGAACGCGTTGGCGCTGATATCGTCCGT 180

Qy         61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsn 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GTATCCGTACCGACGATGGACGCGGCAGAAGCGTTCAAACTCATCAAACAGCAGGTTAAC 240

Qy         81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTGCCGCTGGTGGCTGACATCCACTTCGACTATCGCATTGCGCTGAAAGTAGCGGAATAC 300

Qy        101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMet 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGCGTCGATTGTCTGCGTATTAACCCTGGCAATATCGGTAATGAAGAGCGTATTCGCATG 360

Qy        121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GTGGTTGACTGTGCGCGCGATAAAAACATTCCGATCCGTATTGGCGTTAACGCCGGATCG 420

Qy        141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSer 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTGGAAAAAGATCTGCAAGAAAAGTATGGCGAACCGACGCCGCAGGCGTTGCTGGAATCT 480

Qy        161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GCCATGCGTCATGTTGATCATCTCGATCGCCTGAACTTCGATCAGTTCAAAGTCAGCGTG 540

Qy        181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AAAGCGTCTGACGTCTTCCTCGCTGTTGAGTCTTATCGTTTGCTGGCAAAACAGATCGAT 600

Qy        201 GlnProLeuHisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSer 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CAGCCGTTGCATCTGGGGATCACCGAAGCCGGTGGTGCGCGCAGCGGGGCAGTAAAATCC 660

Qy        221 AlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GCCATTGGTTTAGGTCTGCTGCTGTCTGAAGGCATCGGCGACACGCTGCGCGTATCGCTG 720

Qy        241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GCGGCCGATCCGGTCGAAGAGATCAAAGTCGGTTTCGATATTTTGAAATCGCTGCGTATC 780

Qy        261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspVal 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CGTTCGCGAGGGATCAACTTCATCGCCTGCCCGACCTGTTCGCGTCAGGAATTTGATGTT 840

Qy        281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 ATCGGTACGGTTAACGCGCTGGAGCAACGCCTGGAAGATATCATCACTCCGATGGACGTT 900

Qy        301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TCGATTATCGGCTGCGTGGTGAATGGCCCAGGTGAGGCGCTGGTTTCTACACTCGGCGTC 960

Qy        321 ThrGlyGlyAsnLysLysSerGlyLeuTyrGluAspGlyValArgLysAspArgLeuAsp 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ACCGGCGGCAACAAGAAAAGCGGCCTCTATGAAGATGGCGTGCGCAAAGACCGTCTGGAC 1020

Qy        341 AsnAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AACAACGATATGATCGACCAGCTGGAAGCACGCATTCGTGCGAAAGCCAGTCAGCTGGAC 1080

Qy        361 GluAlaArgArgIleAspValGlnGlnValGluLys 372
              ||||||||||||||||||||||||||||||||||||
Db       1081 GAAGCGCGTCGAATTGACGTTCAGCAGGTTGAAAAA 1116

RESULT 3
AAC82653
ID   AAC82653 standard; DNA; 1119 BP.
XX
AC   AAC82653;
XX
DT   15-MAR-2001  (first entry)
XX
DE   E. coli gcpE DNA.
XX
KW   Isoprenoid; gcpE; yfgB; antimicrobial; transgenic plant; agriculture;
KW   antimycotic; antiparasitic; antiviral; fungicidal; herbicidal; ds.
XX
OS   Escherichia coli.
XX
PN   WO200072022-A1.
XX
PD   30-NOV-2000.
XX
PF   20-MAY-2000; 2000WO-EP04592.
XX
PR   21-MAY-1999;   99DE-1023567.
PR   21-MAY-1999;   99DE-1023568.
XX
PA   (JOMA/) JOMAA H.
XX
PI   Jomaa H;
```

```
XX
DR   WPI; 2001-025196/03.
DR   P-PSDB; AAB45692.
XX
PT   Incorporating gcpE and yfgB genes into viruses and cells, for
PT   increasing isoprenoid content and identifying e.g. antimicrobial
PT   agents, comprises using DNA sequences from bacteria or parasites -
XX
PS   Claim 3; Page 13-15; 36pp; German.
XX
CC   This invention describes a novel method for incorporating gcpE and yfgB
CC   genes into viruses and cells for increasing isoprenoid content and
CC   identifying e.g. antimicrobial agents, comprises using DNA sequences (I)
CC   from the gcpE or yfgB genes of bacteria or parasites or DNA sequences
CC   (II) which hybridize to the specified genes or encode a plastid protein
CC   with the same biological activity as those encoded by the genes. The
CC   invention also describes (1) plant cells containing (I) or (II); (2)
CC   transformed plant cells, and transgenic plants regenerated from them,
CC   that contain (I) or (II); (3) determining the enzymatic activity of a
CC   gcpE protein; or (4) screening compounds (A) that have antimycotic,
CC   antiparasitic or antiviral activity in humans or animals or antiviral,
CC   antiparasitic, fungicidal or herbicidal activity in plants. (I) and (II)
CC   are used: (i) to increase the isoprenoid levels in viruses and cells;
CC   (ii) for determining the enzymatic activity of gcpE and yfgB proteins;
CC   and (iii) to identify compounds that inhibit activity of gcpE, i.e.
CC   potential antibacterial, antimycotic, antiparasitic or antiviral agents
CC   for use in humans or animals, or antiviral, antiparasitic, fungicidal
CC   or herbicidal agents for agriculture.
XX
SQ   Sequence 1119 BP; 272 A; 279 C; 317 G; 251 T; 0 other;

Alignment Scores:
Pred. No.:              7.12e-184       Length:         1119
Score:                  1866.00         Matches:        372
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     22              Gaps:           0

US-09-921-992-78 (1-372) x AAC82653 (1-1119)

Qy          1 MetHisAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCATAACCAGGCTCCAATTCAACGTAGAAAATCAACACGTATTTACGTTGGGAATGTG 60

Qy         21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrThrAsp 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCGATTGGCGATGGTGCTCCCATCGCCGTACAGTCCATGACCAATACGCGTACGACAGAC 120

Qy         41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GTCGAAGCAACGGTCAATCAAATCAAGGCGCTGGAACGCGTTGGCGCTGATATCGTCCGT 180

Qy         61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsn 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GTATCCGTACCGACGATGGACGCGGCAGAAGCGTTCAAACTCATCAAACAGCAGGTTAAC 240

Qy         81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTGCCGCTGGTGGCTGACATCCACTTCGACTATCGCATTGCGCTGAAAGTAGCGGAATAC 300

Qy        101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMet 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGCGTCGATTGTCTGCGTATTAACCCTGGCAATATCGGTAATGAAGAGCGTATTCGCATG 360

Qy        121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GTGGTTGACTGTGCGCGCGATAAAAACATTCCGATCCGTATTGGCGTTAACGCCGGATCG 420

Qy        141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSer 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTGGAAAAAGATCTGCAAGAAAAGTATGGCGAACCGACGCCGCAGGCGTTGCTGGAATCT 480

Qy        161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GCCATGCGTCATGTTGATCATCTCGATCGCCTGAACTTCGATCAGTTCAAAGTCAGCGTG 540

Qy        181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AAAGCGTCTGACGTCTTCCTCGCTGTTGAGTCTTATCGTTTGCTGGCAAAACAGATCGAT 600

Qy        201 GlnProLeuHisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSer 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CAGCCGTTGCATCTGGGGATCACCGAAGCCGGTGGTGCGCGCAGCGGGGCAGTAAAATCC 660

Qy        221 AlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GCCATTGGTTTAGGTCTGCTGCTGTCTGAAGGCATCGGCGACACGCTGCGCGTATCGCTG 720

Qy        241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GCGGCCGATCCGGTCGAAGAGATCAAAGTCGGTTTCGATATTTTGAAATCGCTGCGTATC 780

Qy        261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspVal 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CGTTCGCGAGGGATCAACTTCATCGCCTGCCCGACCTGTTCGCGTCAGGAATTTGATGTT 840

Qy        281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 ATCGGTACGGTTAACGCGCTGGAGCAACGCCTGGAAGATATCATCACTCCGATGGACGTT 900

Qy        301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TCGATTATCGGCTGCGTGGTGAATGGCCCAGGTGAGGCGCTGGTTTCTACACTCGGCGTC 960

Qy        321 ThrGlyGlyAsnLysLysSerGlyLeuTyrGluAspGlyValArgLysAspArgLeuAsp 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ACCGGCGGCAACAAGAAAAGCGGCCTCTATGAAGATGGCGTGCGCAAAGACCGTCTGGAC 1020

Qy        341 AsnAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AACAACGATATGATCGACCAGCTGGAAGCACGCATTCGTGCGAAAGCCAGTCAGCTGGAC 1080

Qy        361 GluAlaArgArgIleAspValGlnGlnValGluLys 372
              ||||||||||||||||||||||||||||||||||||
Db       1081 GAAGCGCGTCGAATTGACGTTCAGCAGGTTGAAAAA 1116

RESULT 4
AAD31203
ID   AAD31203 standard; DNA; 1119 BP.
XX
AC   AAD31203;
XX
DT   31-MAY-2002  (first entry)
XX
DE   Escherichia coli gcpE gene.
XX
KW   gcpE gene; methyl-D-erythritol phosphate; MEP; dimethylallyl diphosphate;
KW   transgenic plant; isoprenoid compound; tocopherol; isopentyl diphosphate;
KW   food; feed source; transfection; single nucleotide polymorphism; SNP;
KW   oxidative stress tolerance; UV tolerance; transformation; plant; ds.
XX
OS   Escherichia coli.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1119
FT                   /*tag=  a
FT                   /product= "Escherichia coli GCPE protein"
XX
PN   WO200212478-A2.
XX
PD   14-FEB-2002.
XX
PF   06-AUG-2001; 2001WO-US24335.
XX
PR   07-AUG-2000; 2000US-223483P.
```

```
XX
PA   (MONS ) MONSANTO TECHNOLOGY LLC.
XX
PI   Boronat A,  Campos N,  Rodriguez-concepcion M,  Rohmer M,  Seeman M;
PI   Valentin HE,  Venkatesh TV,  Venkatramesh M;
XX
DR   WPI; 2002-227151/28.
DR   P-PSDB; AAE19653.
XX
PT   gcpE nucleic acid which is an essential gene of the methyl-D-erythritol
PT   phosphate pathway, encoding a fully defined GCPE protein which is
PT   useful for increasing levels of tocopherol substrates in plants  -
XX
PS   Claim 3; Page 117-119; 155pp; English.
XX
CC   The invention relates to gcpE nucleic acid molecule, an essential gene
CC   of methyl-D-erythritol phosphate (MEP) pathway) that encodes rice,
CC   Arabidopsis thaliana or Escherichia coli GCPE protein. gcpE is useful
CC   for producing a transgenic plant such as Brassica campestris, B. napus,
CC   canola, castor bean, coconut, cotton, crambe, linseed, maize, mustard,
CC   oil palm, peanut, rapeseed, rice, safflower, sesame, soybean, sunflower,
CC   or wheat with an increased isoprenoid (tocopherol) compound level. The
CC   expression of GCPE protein in organisms increases the level of
CC   tocopherol substrate such as isopentyl diphosphate and dimethylallyl
CC   diphosphate biosynthesis. Transgenic organisms overexpressing GCPE
CC   protein can nutritionally enhance food and feed sources. Overexpression
CC   of GCPE protein in transgenic plant may provide tolerance to stresses
CC   e.g., oxidative stress tolerance such as to oxygen or ozone, UV
CC   tolerance, etc. gcpE may be used to obtain nucleic acid molecules from
CC   the same species, and to obtain nucleic acid homologues. gcpE is also
CC   used as or primers. The recombinant vectors are used in plant
CC   transformation or transfection. gcpE an also act as markers capable of
CC   detecting polymorphisms such as single nucleotide polymorphisms (SNPs).
CC   gcpE is also used to determine the level or pattern of expression of
CC   the protein. The present sequence is Escherichia coli gcpE gene.
XX
SQ   Sequence 1119 BP; 272 A; 279 C; 317 G; 251 T; 0 other;

Alignment Scores:
Pred. No.:                7.12e-184      Length:          1119
Score:                    1866.00        Matches:         372
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       24             Gaps:            0

US-09-921-992-78 (1-372) x AAD31203 (1-1119)

Qy          1 MetHisAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCATAACCAGGCTCCAATTCAACGTAGAAAATCAACACGTATTTACGTTGGGAATGTG 60

Qy         21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrThrAsp 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCGATTGGCGATGGTGCTCCCATCGCCGTACAGTCCATGACCAATACGCGTACGACAGAC 120

Qy         41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GTCGAAGCAACGGTCAATCAAATCAAGGCGCTGGAACGCGTTGGCGCTGATATCGTCCGT 180

Qy         61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsn 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GTATCCGTACCGACGATGGACGCGGCAGAAGCGTTCAAACTCATCAAACAGCAGGTTAAC 240

Qy         81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTGCCGCTGGTGGCTGACATCCACTTCGACTATCGCATTGCGCTGAAAGTAGCGGAATAC 300

Qy        101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMet 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGCGTCGATTGTCTGCGTATTAACCCTGGCAATATCGGTAATGAAGAGCGTATTCGCATG 360

Qy        121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GTGGTTGACTGTGCGCGCGATAAAAACATTCCGATCCGTATTGGCGTTAACGCCGGATCG 420

Qy        141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSer 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTGGAAAAAGATCTGCAAGAAAAGTATGGCGAACCGACGCCGCAGGCGTTGCTGGAATCT 480

Qy        161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GCCATGCGTCATGTTGATCATCTCGATCGCCTGAACTTCGATCAGTTCAAAGTCAGCGTG 540

Qy        181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AAAGCGTCTGACGTCTTCCTCGCTGTTGAGTCTTATCGTTTGCTGGCAAAACAGATCGAT 600

Qy        201 GlnProLeuHisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSer 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CAGCCGTTGCATCTGGGGATCACCGAAGCCGGTGGTGCGCGCAGCGGGGCAGTAAAATCC 660

Qy        221 AlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GCCATTGGTTTAGGTCTGCTGCTGTCTGAAGGCATCGGCGACACGCTGCGCGTATCGCTG 720

Qy        241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GCGGCCGATCCGGTCGAAGAGATCAAAGTCGGTTTCGATATTTTGAAATCGCTGCGTATC 780

Qy        261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspVal 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CGTTCGCGAGGGATCAACTTCATCGCCTGCCCGACCTGTTCGCGTCAGGAATTTGATGTT 840

Qy        281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 ATCGGTACGGTTAACGCGCTGGAGCAACGCCTGGAAGATATCATCACTCCGATGGACGTT 900

Qy        301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TCGATTATCGGCTGCGTGGTGAATGGCCCAGGTGAGGCGCTGGTTTCTACACTCGGCGTC 960

Qy        321 ThrGlyGlyAsnLysLysSerGlyLeuTyrGluAspGlyValArgLysAspArgLeuAsp 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ACCGGCGGCAACAAGAAAAGCGGCCTCTATGAAGATGGCGTGCGCAAAGACCGTCTGGAC 1020

Qy        341 AsnAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AACAACGATATGATCGACCAGCTGGAAGCACGCATTCGTGCGAAAGCCAGTCAGCTGGAC 1080

Qy        361 GluAlaArgArgIleAspValGlnGlnValGluLys 372
              ||||||||||||||||||||||||||||||||||||
Db       1081 GAAGCGCGTCGAATTGACGTTCAGCAGGTTGAAAAA 1116

RESULT 5
AAT42063
ID   AAT42063 standard; DNA; 1830121 BP.
XX
AC   AAT42063;
XX
DT   14-SEP-1999  (first entry)
XX
DE   Haemophilus influenzae complete genome sequence.
XX
KW   Genome; bacterium; Haemophilus influenzae; computer readable medium;
KW   expression modulating fragment; regulation; gene expression; vector;
KW   organism; open reading frame; ORF; ds.
XX
OS   Haemophilus influenzae.
XX
PN   WO9633276-A1.
XX
PD   24-OCT-1996.
XX
PF   22-APR-1996;   96WO-US05320.
XX
```

PR 07-JUN-1995; 95US-0487429.
PR 21-APR-1995; 95US-0426787.
PR 07-JUN-1995; 95US-0476102.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (UYJO) UNIV JOHNS HOPKINS.
XX
PI Adams MD, Fleischmann RD, Smith HO, Venter JC, White O;
XX
DR WPI; 1996-485782/48.
XX
PT Haemophilus influenzae Rd genome recorded on computer readable
PT medium - useful for identifying commercially important nucleic acid
PT fragments by homology searching
XX
PS Claim 1; Page 77.2-77.1091; 1291pp; English.
XX
CC This sequence represents the complete genome sequence of the bacterium
CC Haemophilus influenzae strain Rd. The invention relates to a computer
CC readable medium (CRM) having recorded upon it the complete H.influenzae
CC nucleotide sequence (I), a representative fragment of (I) or a nucleotide
CC sequence at least 99% identical to (I). By providing the full-length
CC genomic sequence in a computer readable form, it is possible to identify
CC commercially important nucleic acid fragments and expression modulating
CC fragments (EMFs) of the Haemophilus genome. The EMFs can be used to
CC regulate the expression of a nucleic acid molecule. Vectors and altered
CC organisms comprising the predicted ORFS can be used to produce any of the
CC polypeptide fragments of the H. influenzae Rd genome.
XX
SQ Sequence 1830121 BP; 567399 A; 350615 C; 347389 G; 564036 T; 682 other;

Alignment Scores:

Pred. No.:	2.6e-152	Length:	1830121
Score:	1603.50	Matches:	320
Percent Similarity:	93.66%	Conservative:	20
Best Local Similarity:	88.15%	Mismatches:	22
Query Match:	85.93%	Indels:	1
DB:	17	Gaps:	1

US-09-921-992-78 (1-372) x AAT42063 (1-1830121)

```
Qy          4 GlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnValProIleGly 23
              |||      |||:::||||||:::|||||||:::|||||||||||||||||||||||||
Db     390972 CAGCCAACTATTAAGCGTCGTGAATCGACAAAAATTTATGTGGGAAATGTACCAATTGGT 391031

Qy         24 AspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrThrAspValGluAla 43
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     391032 GGGGATGCGCCTATTGCCGTGCAATCAATGACAAATACTCGCACCACTGATGTGGAAGCG 391091

Qy         44 ThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArgValSerVal 63
              ||||||   ||||||||||:::|||||||||||||||||||||||||||||||||||||
Db     391092 ACAGTTGCTCAAATTAAATCATTAGAACGTGTTGGTGCAGATATTGTTCGTGTATCTGTT 391151

Qy         64 ProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsnValProLeu 83
              ||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db     391152 CCAACAATGGATGCTGCGGAAGCATTTAAACAAATTAAACAACAAGTGAATGTTCCGCTC 391211

Qy         84 ValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAsp 103
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     391212 GTAGCAGATATTCATTTCGACTATCGTATCGCGTTAAAAGTCGCAGAATATGGAGTGGAT 391271

Qy        104 CysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMetValValAsp 123
              ||||||||||||||||||||||||||||   |||:::|||:::|||   |||||||||
Db     391272 TGTTTACGTATCAATCCTGGCAACATTGGTCGTGAAGATCGCGTCCGTGCCGTTGTTGAT 391331

Qy        124 CysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLys 143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     391332 TGTGCGCGAGACAAAAATATTCCGATTCGTATTGGTGTAAATGCAGGCTCTTTAGAAAAA 391391

Qy        144 AspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArg 163
              ||||||||||||||||||||||||||||||:::|||||||||||||||||:::|||
Db     391392 GATTTGCAAGAAAAATATGGCGAACCAACGCCAGAAGCCTTGTTAGAATCCGCATTGCGT 391451

Qy        164 HisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSer 183
              ||||||:::   ||||||||||||||||||||||||||||||||||||||||||||||
Db     391452 CATGTAGAAATTCTAGATCGTCTTAACTTCGATCAGTTTAAAGTGAGCGTAAAAGCCTCC 391511

Qy        184 AspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeu 203
              ||||||||||||||||||||||||||||||||||||||||||   |||   |||||||||
Db     391512 GATGTATTCTTAGCGGTTGAATCTTATCGTTTACTGGCTAAAGCAATTAAACAGCCTTTA 391571

Qy        204 HisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGly 223
              ||||||||||||||||||||||||||||||||:::|||||||||||||||||||:::|||
Db     391572 CATTTAGGCATTACAGAAGCAGGTGGCGCACGCGCTGGTGCAGTAAAATCTGCAGTGGGT 391631

Qy        224 LeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAsp 243
              ||||||:::|||||||:::|||||||||||||||||||||||||||||||||||||||||
Db     391632 TTAGGAATGTTATTAGCTGAGGGCATTGGCGATACACTACGCGTCTCTTTGGCGGCAGAT 391691

Qy        244 ProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArg 263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     391692 CCTGTAGAGGAAATCAAAGTCGGTTTTGATATTTTGAAATCTTTACGGATTCGTTCAAGA 391751

Qy        264 GlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspValIleGlyThr 283
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     391752 GGAATTAACTTTATTGCTTGCCCAACCTGTTCTCGCCAAGAATTTGATGTAATCGGTACA 391811

Qy        284 ValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspValSerIleIle 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     391812 GTAAATGCGCTAGAACAACGCCTTGAAGATATTATTACACCAATGGATGTATCTATTATC 391871

Qy        304 GlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyGly 323
              ||||||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db     391872 GGTTGTGTAGTGAATGGTCCTGGCGAGGCACTCGTCTCCGATCTCGGCGTAACGGGCGGT 391931

Qy        324 AsnLysLysSerGlyLeuTyrGluAspGlyValArg---LysAspArgLeuAspAsnAsn 342
              |||||||||||||||   |||   ||||||   |||   |||:::|||   ||||||
Db     391932 AACAAAAAAAGCGGTTATTATCTTGACGGAGAACGCCAAAAAGAGCGTTTTGATAACGAA 391991

Qy        343 AspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAspGluAla 362
              |||::::::::::|||||||||||||:::|||||||||||||   ::::::   |||
Db     391992 GATATAGTGAACCAATTAGAAGCAAAAATTCGTGCGAAAGTCGCACGACAAGATCCAAAA 392051

Qy        363 ArgArgIle 365
                 ||||||
Db     392052 AACAGAATT 392060
```

RESULT 6
ABA92787
ID ABA92787 standard; DNA; 640681 BP.
XX
AC ABA92787;
XX
DT 27-MAR-2002 (first entry)
XX
DE Buchnera sp. genomic DNA SEQ ID NO:1.
XX
KW Buchnera; cockroach-symbiotic bacterium; cockroach extermination;
KW circular; ds.
XX
OS Buchnera sp.
XX
PN JP2001292771-A.
XX
PD 23-OCT-2001.
XX
PF 07-APR-2000; 2000JP-0107160.
XX
PR 07-APR-2000; 2000JP-0107160.
XX
PA (RIKA) RIKAGAKU KENKYUSHO.
XX
DR WPI; 2002-126043/17.
XX
PT A genomic DNA of cockroach-symbiotic bacterium -
XX

```
PS   Claim 1; Page 16-230; 237pp; Japanese.
XX
CC   The present invention describes a gene (I) derived from Buchnera sp.
CC   containing the DNA (a) or (b), (a) has a fully defined base pair
CC   sequence selected from a table of sequences found in the Buchnera sp.
CC   genomic DNA of ABA92787 given in the specification or is a DNA selected
CC   from complementary DNA sequences, and (b) is a DNA which hybridises with
CC   the DNA (a) and encodes a protein. Also described are: (1) a recombinant
CC   vector (II) containing (I); (2) a transformant (III) containing (II);
CC   (3) a genomic DNA of Buchnera sp. containing the sequence given in
CC   ABA92787; (4) a plasmid derived from Buchnera sp. containing DNA (c) or
CC   (d), (c) is a DNA containing a fully defined sequence given in ABA92788
CC   or ABB92789 and (d) is a plasmid which hybridises with a DNA; and (5) a
CC   method for the preparation of a protein in which (III) is cultured and
CC   the expression protein of the objective protein is collected from the
CC   resultant culture. The DNA is useful for developing agricultural
CC   chemicals for exterminating cockroaches. The present sequence represents
CC   the specifically claimed Buchnera sp. genomic DNA sequence, from the
CC   present invention.
XX
SQ   Sequence 640681 BP; 237522 A; 83822 C; 84757 G; 234580 T; 0 other;

Alignment Scores:
Pred. No.:                1.41e-119      Length:          640681
Score:                    1282.00        Matches:         242
Percent Similarity:       83.88%         Conservative:    65
Best Local Similarity:    66.12%         Mismatches:      57
Query Match:              68.70%         Indels:          2
DB:                       24             Gaps:            2

US-09-921-992-78 (1-372) x ABA92787 (1-640681)

Qy          1 MetHisAsnGlnAlaPro---IleGlnArgArgLysSerThrArgIleTyrValGlyAsn 19
              |||:::|||:::          |||   ||||||||||||   ||||||||||||||||
Db     314272 ATGAATAATAAGTGTAAAATTATCAATAGAAGAAAATCTGATCGTATTTATGTTGGAAAA 314331

Qy         20 ValProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrThr 39
              |||   ||||||:::   ||||||||||:::||||||||||||||||||||||||||||||
Db     314332 GTCGCTATTGGCAATAATGCGCCAATATCAGTTCAATCTATGACAAATACTCGTACTACT 314391

Qy         40 AspValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleVal 59
              ::::::      |||:::||||||||||      |||::::::::||||||   |||||||||
Db     314392 AATATCTCTGAAACTATTAATCAAATCTTAGAGTTACAAAAAGTAGGAGTAGATATTGTT 314451

Qy         60 ArgValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnVal 79
              |||:::|||:::|||   :::   ||||||||||:::||||||   |||||||:::|||
Db     314452 CGTATTTCTATACCCAATTTAAAAGCTGCAGAATCATTCAAAGAAATAAAAAAACAAACA 314511

Qy         80 AsnValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGlu 99
              ||||||||||||:::||||||||||||||||||||||||||:::|||||||:::      :::
Db     314512 AATGTTCCATTGATTGCAGATATACATTTTGATTACAGATTAGCTTTACAAGCTATAAAA 314571

Qy        100 TyrGlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArg 119
              ||||||   ||||||||||||||||||||||||||||||||||||:::   |||:::
Db     314572 TATGGTGCAGATTGTTTAAGAATTAATCCTGGGAATATTGGAAATAAAAGAAGAGTATCA 314631

Qy        120 MetValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGly 139
                 ::::::      |||:::|||:::||||||||||||||||||||||||||||||||
Db     314632 GAAATCATTTCTTACGCAAAAGATGAAAATATTCCAATTCGTATTGGTGTTAATGCTGGA 314691

Qy        140 SerLeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuGluGlu 159
              ||||||||||||||:::   :::||||||      |||||||||   |||||||:::|||
Db     314692 TCTTTAGAAAAAGATATATTAAAAAAATATAAAATACCTACTCCAGATGCATTAGTAGAA 314751

Qy        160 SerAlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSer 179
              ||||||||||||||:::::::::   |||   ||||||||||:::||||||||||||||||
Db     314752 TCAGCTATGAGGCATATTGAATACTTTGATGCTTTAAATTTTAATCAATTTAAAGTTAGT 314811

Qy        180 ValLysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIle 199
              ||||||||||||||||||||||||||:::||||||||||||:::|||   |||:::|||
Db     314812 GTTAAAGCGTCTGATGTATTTTTAGCTATTGAATCATATCGCATGTTAGGAAAAAAAATT 314871

Qy        200 AspGlnProLeuHisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLys 219
                 ||||||||||||:::||||||||||||:::||||||   |||:::|||   ||||||
Db     314872 ACACAACCTTTGCATATTGGCATAACTGAATCCGGTGGTTTAAGGAATGGAACAGTTAAA 314931

Qy        220 SerAlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSer 239
              |||:::||||||:::   |||||||||   ||||||||||||||||||:::|||||||||
Db     314932 TCATCTATAGGTATTGCTTTATTATTATTAGAAGGCATTGGAGATACAATACGAGTTTCA 314991

Qy        240 LeuAlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArg 259
              |||||||||   |||   ||||||:::|||||||||:::||||||||||||   |||
Db     314992 TTAGCGGCACATCCAACTGAAGAAGTAAAAGTAGGTTATGACATTTTAAAAGTTTTATCT 315051

Qy        260 IleArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAsp 279
              :::|||:::||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     315052 TTAAGAGCAAGAGGTATTAATTTTATTGCTTGTCCTACTTGTTCTAGACAAGAATTTGAT 315111

Qy        280 ValIleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAsp 299
              ||||||   |||||||||   |||||||:::   ||||||||||||   ||||||:::|||
Db     315112 GTAATTAATACAGTAAATCAACTAGAAAAAAATCTAGAAGATATCTCGACTCCCATAGAT 315171

Qy        300 ValSerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGly 319
              ||||||||||||||||||||||||||||||   ||||||||:::   ::::::|||||||||
Db     315172 GTATCAATTATTGGTTGCGTTGTTAATGGAATAGGTGAATCTAAAATAGCAACTTTAGGT 315231

Qy        320 ValThrGlyGlyAsnLysLysSerGlyLeuTyrGluAspGlyValArg---LysAspArg 338
              :::   |||   :::||||||||||      ||||||||||||||||||   |||::::::
Db     315232 CTAGCAGGAAGTCATAAGAAAAGTGCATTTTATGAAGACGGAGTAAGACAAAAAGAAAAA 315291

Qy        339 LeuAspAsnAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGln 358
              :::   |||   :::::::|||:::::::::|||   :::|||||   |||      :::
Db     315292 ATAAAAAACGAAGAAATTATAGAAAAAATGGAAATTAAAATTCGAAAAAAAATAGACAAA 315351

Qy        359 LeuAspGluAlaArgArg 364
              |||:::   :::::::::
Db     315352 TTAAATAATTCAAAAAAA 315369

RESULT 7
AAF28536
ID   AAF28536 standard; DNA; 33140 BP.
XX
AC   AAF28536;
XX
DT   04-APR-2001  (first entry)
XX
DE   Genomic fragment #23.
XX
KW   Genomic library; bacteria; human upper airway; otitis media; sinusitis;
KW   bronchopulmonary; endocarditis; meningitis; ss.
XX
OS   Moraxella catarrhalis.
XX
PN   WO200078968-A2.
XX
PD   28-DEC-2000.
XX
PF   16-JUN-2000; 2000WO-US16649.
XX
PR   18-JUN-1999;   99US-0140121.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Lagace RE,  Patterson C,  Berg KL;
XX
DR   WPI; 2001-041427/05.
XX
PT   Genomic library for identifying diagnostic and therapeutic
PT   compositions, and for identifying virulence factors, regulatory
PT   elements and drug targets, comprises Moraxella catarrhalis nucleic
PT   acids -
XX
PS   Claim 1; Page 191-199; 545pp; English.
XX
CC   The present invention relates to a Moraxella catarrhalis genomic library
CC   comprising of a combination of 41 nucleic acid molecules (see
```

```
CC   AAF28514-AAF28554). The library has a number of uses described in the
CC   specification e.g. is useful for aidentifying diagnostic and therapeutic
CC   compositions. M. catarrhalis (Branhamella catarrhalis) is a large
CC   aerobic, gram-negative diplococcus, normally found among the bacterial
CC   flora of human upper airways. M. catarrhalis is known to cause acute,
CC   localised infections such as otitis media, sinusitis and bronchopulmonary
CC   infection and life-threatening, systemic diseases including endocarditis
CC   and meningitis.
XX
SQ   Sequence 33140 BP; 9472 A; 6496 C; 7384 G; 9787 T; 1 other;

Alignment Scores:
Pred. No.:                2.81e-109      Length:          33140
Score:                    1166.00        Matches:         220
Percent Similarity:       78.92%         Conservative:    72
Best Local Similarity:    59.46%         Mismatches:      76
Query Match:              62.49%         Indels:          2
DB:                       22             Gaps:            1

US-09-921-992-78 (1-372) x AAF28536 (1-33140)

Qy          1 MetHisAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnVal 20
              :::||||||      ||||||:::||||||   :::   :::||||||||||||||||||
Db      21741 ATACACAAT------CGCATTAAACGCCGCCCTACCAAAAAAATCTATGTTGGCAATGTC 21794

Qy         21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrThrAsp 40
              ||||||      ||||||||:::|||||||||||||||||||||   |||   |||
Db      21795 GCCATCGGCGGCGATGCACCCATCAGCGTGCAAAGTATGACAAATACAAACACCTGTGAT 21854

Qy         41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60
              ::::::||||||||| ||||||||::                ||||||||||::::::|||
Db      21855 ATTGATGCAACAGTTGCACAAATTCAGCGATGCGTGGACGCAGGTGCGGACTTGATGCGT 21914

Qy         61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsn 80
              ||||||   ||||||||||::::::      ||||||      ||||||:::::::|||:::
Db      21915 GTATCGACACCAACGATGGAGAGCGTCGCTGCCTTTGCTGAGATAAAAAAACGCGTATCC 21974

Qy         81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100
              :::||||||:::||||||||||||||||||:::::||||||:::   ||||||:::
Db      21975 ATTCCACTCATTGCCGATATTCATTTTGATCACAAAATCGCCATTGCAGTGGCGGATGTG 22034

Qy        101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMet 120
              |||   ||||||||||||||||||||||||||||||||||:::::::::::::::
Db      22035 GGTGCAGACTGCCTGCGTATCAATCCAGGTAATATTGGCAATGATCAAAAAGTCAAAGAA 22094

Qy        121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140
              ||||||      ||||||      |||:::||||||||||||||||||||||||||||||
Db      22095 GTGGTTGCAGCGGCACGCCATCATAATGTGCCGATTCGTATCGGTGTGAATGCAGGCTCA 22154

Qy        141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSer 160
              ||||||||||||||||||:::|||||||   |||||||||   :::|||:::||||||||||
Db      22155 CTTGAAAAGGATTTACAAAAAAAATATCGAGAACCGACAGGTGAGGCGATGCTAGAATCG 22214

Qy        161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180
              |||:::|||||||:::||||   |||:::   |||||||||      :::|||:::||||||
Db      22215 GCGTTACGCCACATTGATATTTTAGAAAATCTTAATTTTCAGAATTATAAAATCTCAGTC 22274

Qy        181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200
              |||||||||:::|||||||||   :::::::::||||||||||:::::   ||||||||
Db      22275 AAAGCCTCCAATGTATTTTTAACCCTTGATGCGTATCGATTGATTTCAGCACAAATTGAT 22334

Qy        201 GlnProLeuHisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSer 220
              ||||||||||||||||:::||||||||||||      |||:::|||   ||||||||
Db      22335 AATCCATTGCATCTTGGCGTTACCGAAGCGGGCGTATATCGCACTGGCACGGTCAAATCT 22394

Qy        221 AlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240
              ||||||   ||||||  ||||||   :::||||||||||||||:::|||:::||||||
Db      22395 GCCATTGCTTTGGGCGGTTTGCTACTTGATGGCATTGGCGATACCATGCGTATCTCTTTG 22454

Qy        241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260
              ||||||   |||   ||||||||||||:::|||||||||||||||||||||   |||
Db      22455 GCGGCAGCACCAGAAGAAGAGATTAAAATTGGCTTTGACATTCTAAAATCACTGGGTATT 22514

Qy        261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspVal 280
              ||||||   |||:::|||||||||||||||||||:::|||||||||||||||||||||||
Db      22515 CGTAGTAATGGCGTCAATTTTATTGCTTGTCCAAGTTGTAGCCGTCAAGAGTTTGATGTG 22574

Qy        281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300
              |||      :::|||   ||||||   ||||||||||||||      |||:::|||:::
Db      22575 ATCAAAGTGATGAATGAGCTTGAAGCTCGCTTAGAGGATATTCGTGAGCCACTAGATCTG 22634

Qy        301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320
              |||:::|||||||||   ||||||||||||||||||||      :::|||:::||||||
Db      22635 TCAGTTATTGGCTGTAAAGTCAATGGGCCAGGTGAAGCCAAAGAAGCGACGATCGGTGTC 22694

Qy        321 ThrGlyGlyAsnLysLysSerGlyLeuTyrGluAspGlyValArgLysAspArgLeuAsp 340
              |||   :::      |||   :::|||:::   |||   :::      :::|||
Db      22695 GTGGGAGCCAGCCCCAATAGTTTGGTTTATAAACATGGTGCAAAAAGCCATCTGATTGAT 22754

Qy        341 AsnAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360
                    :::::::|||:::::::||||||   :::|||||||:::         :::
Db      22755 ACCAAAAAATTGGTAGATGAGATTGAGGCGATGGTGCGTGCTCAAGTCAAAGACATTCAA 22814

Qy        361 GluAlaArgArgIleAspValGlnGlnVal 370
              :::   |||      ::::::   ::::::
Db      22815 GATAAACGGGCAAATGAGATTATCCGTATA 22844

RESULT 8
ABQ67966
ID   ABQ67966 standard; DNA; 1107 BP.
XX
AC   ABQ67966;
XX
DT   29-AUG-2002  (first entry)
XX
DE   Listeria monocytogenes EGD DNA sequence #90.
XX
KW   Antibacterial; Listeria; food contamination; mutational analysis;
KW   infection; ds.
XX
OS   Listeria monocytogenes EGD.
XX
PN   WO200228891-A2.
XX
PD   11-APR-2002.
XX
PF   04-OCT-2001; 2001WO-FR03061.
XX
PR   04-OCT-2000; 2000FR-0012697.
XX
PA   (INSP ) INST PASTEUR.
PA   (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI   Kunst F,  Glaser P;
XX
DR   WPI; 2002-332479/37.
XX
PT   New genomic sequences from Listeria species, useful for detection,
PT   treatment and prevention of infection, also related polypeptides,
PT   antibodies and modulators  -
XX
PS   Claim 16; SEQ ID 779; 180pp; French.
XX
CC   The present invention relates to nucleic acid sequences
CC   (ABQ67188-ABQ71212) from Listeria sp. The sequences are useful as probes
CC   and primers for identification and/or detection of Listeria (e.g. as
CC   contaminants in foods, or mutational analysis) and for analysis of
CC   gene expression. Proteins encoded by the nucleic acid sequences can be
CC   used to screen for compounds that modulate gene expression, replication
CC   and pathogenicity of Listeria (potential therapeutic agents), also for
CC   treating infections by Listeria, and are useful as immunogens in
CC   anti-Listeria vaccines.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic format
CC   directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
```

SQ Sequence 1107 BP; 348 A; 178 C; 255 G; 326 T; 0 other;

Alignment Scores:

Pred. No.:	2.34e-80	Length:	1107
Score:	868.00	Matches:	178
Percent Similarity:	67.68%	Conservative:	67
Best Local Similarity:	49.17%	Mismatches:	115
Query Match:	46.52%	Indels:	2
DB:	24	Gaps:	2

US-09-921-992-78 (1-372) x ABQ67966 (1-1107)

```
Qy         10 ArgLysSerThrArg---IleTyrValGlyAsnValProIleGlyAspGlyAlaProIle 28
              |||:::::|||||   :::   |||||||||:::   ||||||             :::
Db         19 CGCGAAAACACTCGCCCAGTCCAAGTGGGTAATTTAACTATTGGTGGTAGTGAGGAATTA 78

Qy         29 AlaValGlnSerMetThrAsnThrArgThrThrAspValGluAlaThrValAsnGlnIle 48
                 :::|||||||||||||   |||:::|||   ||||||||||||||||||   :::|||
Db         79 ACTATCCAAAGCATGACTACTACAAAGACACATGATGTCGAAGCAACAGTAGCAGAAATT 138

Qy         49 LysAlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMetAspAla 68
                 ||||||     |||      ||||||||||||:::   |||             |||
Db        139 CACCGATTAGAAGAAGCTGGTTGTCAGATTGTGCGAGTTGCTTGTCCTGATGAACGTGCA 198

Qy         69 AlaGluAlaPheLysLeuIleLysGlnGlnValAsnValProLeuValAlaAspIleHis 88
              |||   |||        ||||||:::::::::::::::||||||||||||||||||||
Db        199 GCGAATGCTCTTAGTGCCATCAAGAAAAAGATTCATATTCCGCTTGTGGCAGATATTCAT 258

Qy         89 PheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArgIleAsn 108
              ||||||||||||:::|||||||||      :::   |||||||||   :::|||||||||
Db        259 TTTGATTACCGACTAGCACTTAAAGCTATTGATGCAGGTGTTGACAAAATTCGGATTAAT 318

Qy        109 ProGlyAsnIleGlyAsnGluGluArgIleArgMetValValAspCysAlaArgAspLys 128
              ||||||||||||||      :::|||:::      ||||||:::   |||:::   |||
Db        319 CCTGGTAACATTGGTCGCCGTGATCGGGTGGAAAAAGTGGTTAATGCTGCTAAAGCAAAA 378

Qy        129 AsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGlnGluLys 148
              ||||||||||||||||||||||||||||||||||||||||||||||||   :::   :::|||
Db        379 AATATTCCAATCCGTATTGGGGTTAATGCTGGTAGTTTAGAAAAGAAAATTATTCAAAAA 438

Qy        149 TyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeu 168
              ||||||   ||||||         ::::::|||||||||:::   |||:::      |||
Db        439 TATGGTTACCCTACTGCTGACGGAATGGTAGAAAGTGCACTTGCCCATATTAAAATTCTC 498

Qy        169 AspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAla 188
              :::   |||:::|||                ||||||:::|||||||||||||||   ||||||
Db        499 GAAGATTTAGATTTTTATGATATCATTGTTTCTTTGAAGGCTTCTGATGTGAATTTAGCA 558

Qy        189 ValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeuGlyIleThr 208
              :::|||:::|||         ::::::      :::   ||||||||||||||||||||
Db        559 ATTGAAGCTTATGATAAAGCTAGTCGCGCATTTAATTATCCTCTGCATCTCGGAATTACA 618

Qy        209 GluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeuLeuLeu 228
              |||:::|||         :::|||   :::|||||||||   |||||||||   :::|||
Db        619 GAATCTGGTACACAATTTGCTGGAGGAATAAAAAGTGCTGCTGGTTTAGGAGCGATACTC 678

Qy        229 SerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGluGluIle 248
              |||   ||||||||||:::||||||||||||||||||:::||||||||||||||||||||
Db        679 AGTTTGGGCATTGGAAATACATTACGAGTATCTTTGAGTGCTGATCCTGTGGAAGAAATA 738

Qy        249 LysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsnPheIle 268
              ||||||      ::::::|||||||||      :::   |||             |||
Db        739 AAAGTGGCCCGGGAAGTTTTAAAATCATTTGGTCTTTCCTCGAATGCCGCAATGCTTATC 798

Qy        269 AlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsnAlaLeuGlu 288
              :::||||||||||||   |||   |||   |||:::|||         |||   :::|||
Db        799 TCCTGCCCTACTTGCGGTCGAATAGAGATTGATTTAATTCGTATCGCTAATGAAGTGGAA 858

Qy        289 GlnArgLeuGluAspIleIleThrProMetAspValSerIleIleGlyCysValValAsn 308
                 :::      |||      |||:::   |||::::::::::|||||||   ||||||
Db        859 AATTACATAGCAAAGATTGAAGTTCCGATTAAAGTAGCCGTGCTTGGCTGTGCGGTCAAC 918

Qy        309 GlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyGlyAsnLysLysSerGly 328
              ||||||||||||||||      :::   :::|||:::   |||   |||   :::
Db        919 GGCCCTGGAGAAGCTCGCGAAGCCGATATCGGAATTGCTGGTTCAAACGGAGAAGGCCTT 978

Qy        329 LeuTyrGluAspGlyValArgLysAspArgLeuAspAsnAsnAspMetIleAspGlnLeu 348
              |||:::      |||          ::::::          ||||||||||:::|||
Db        979 CTTTTTAGACATGGTAAAATTATTCGAAAAGTACCGGAAGCTATTATGATAGACGAACTT 1038

Qy        349 GluAlaArgIleArgAlaLysAlaSerGlnLeuAspGluAlaArgArgIleAspValGln 368
              :::      |||         |||   :::             ::::::||||||:::::
Db       1039 AAGAAAGAAATTGATATTTTGGCAGAAGAATTT---TTTGTGAAGAAAATAGATTTGGAA 1095

Qy        369 GlnVal 370
                 :::
Db       1096 AGCCTT 1101
```

RESULT 9
ABQ69993

```
ID   ABQ69993 standard; DNA; 1107 BP.
XX
AC   ABQ69993;
XX
DT   29-AUG-2002  (first entry)
XX
DE   Listeria monocytogenes EGDe DNA sequence #205.
XX
KW   Antibacterial; Listeria; food contamination; mutational analysis;
KW   infection; ds.
XX
OS   Listeria monocytogenes EGDe.
XX
PN   WO200228891-A2.
XX
PD   11-APR-2002.
XX
PF   04-OCT-2001; 2001WO-FR03061.
XX
PR   04-OCT-2000; 2000FR-0012697.
XX
PA   (INSP ) INST PASTEUR.
PA   (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI   Kunst F,  Glaser P;
XX
DR   WPI; 2002-332479/37.
XX
PT   New genomic sequences from Listeria species, useful for detection,
PT   treatment and prevention of infection, also related polypeptides,
PT   antibodies and modulators -
XX
PS   Claim 16; SEQ ID 2806; 180pp; French.
XX
CC   The present invention relates to nucleic acid sequences
CC   (ABQ67188-ABQ71212) from Listeria sp. The sequences are useful as probes
CC   and primers for identification and/or detection of Listeria (e.g. as
CC   contaminants in foods, or mutational analysis) and for analysis of
CC   gene expression. Proteins encoded by the nucleic acid sequences can be
CC   used to screen for compounds that modulate gene expression, replication
CC   and pathogenicity of Listeria (potential therapeutic agents), also for
CC   treating infections by Listeria, and are useful as immunogens in
CC   anti-Listeria vaccines.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic format
CC   directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1107 BP; 348 A; 178 C; 255 G; 326 T; 0 other;
```

Alignment Scores:

Pred. No.:	2.34e-80	Length:	1107
Score:	868.00	Matches:	178
Percent Similarity:	67.68%	Conservative:	67
Best Local Similarity:	49.17%	Mismatches:	115
Query Match:	46.52%	Indels:	2
DB:	24	Gaps:	2

US-09-921-992-78 (1-372) x ABQ69993 (1-1107)

```
Qy         10 ArgLysSerThrArg---IleTyrValGlyAsnValProIleGlyAspGlyAlaProIle 28
Db         19 CGCGAAAACACTCGCCCAGTCCAAGTGGGTAATTTAACTATTGGTGGTAGTGAGGAATTA 78

Qy         29 AlaValGlnSerMetThrAsnThrArgThrThrAspValGluAlaThrValAsnGlnIle 48
Db         79 ACTATCCAAAGCATGACTACTACAAAGACACATGATGTCGAAGCAACAGTAGCAGAAATT 138

Qy         49 LysAlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMetAspAla 68
Db        139 CACCGATTAGAAGAAGCTGGTTGTCAGATTGTGCGAGTTGCTTGTCCTGATGAACGTGCA 198

Qy         69 AlaGluAlaPheLysLeuIleLysGlnGlnValAsnValProLeuValAlaAspIleHis 88
Db        199 GCGAATGCTCTTAGTGCCATCAAGAAAAAGATTCATATTCCGCTTGTGGCAGATATTCAT 258

Qy         89 PheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArgIleAsn 108
Db        259 TTTGATTACCGACTAGCACTTAAAGCTATTGATGCAGGTGTTGACAAAATTCGGATTAAT 318

Qy        109 ProGlyAsnIleGlyAsnGluGluArgIleArgMetValValAspCysAlaArgAspLys 128
Db        319 CCTGGTAACATTGGTCGCCGTGATCGGGTGGAAAAAGTGGTTAATGCTGCTAAAGCAAAA 378

Qy        129 AsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGlnGluLys 148
Db        379 AATATTCCAATCCGTATTGGGGTTAATGCTGGTAGTTTAGAAAAGAAAATTATTCAAAAA 438

Qy        149 TyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeu 168
Db        439 TATGGTTACCCTACTGCTGACGGAATGGTAGAAAGTGCACTTGCCCATATTAAAATTCTC 498

Qy        169 AspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAla 188
Db        499 GAAGATTTAGATTTTTATGATATCATTGTTTCTTTGAAGGCTTCTGATGTGAATTTAGCA 558

Qy        189 ValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeuGlyIleThr 208
Db        559 ATTGAAGCTTATGATAAAGCTAGTCGCGCATTTAATTATCCTCTGCATCTCGGAATTACA 618

Qy        209 GluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeuLeuLeu 228
Db        619 GAATCTGGTACACAATTTGCTGGAGGAATAAAAAGTGCTGCTGGTTTAGGAGCGATACTC 678

Qy        229 SerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGluGluIle 248
Db        679 AGTTTGGGCATTGGAAATACATTACGAGTATCTTTGAGTGCTGATCCTGTGGAAGAAATA 738

Qy        249 LysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsnPheIle 268
Db        739 AAAGTGGCCCGGGAAGTTTTAAAATCATTTGGTCTTTCCTCGAATGCCGCAATGCTTATC 798

Qy        269 AlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsnAlaLeuGlu 288
Db        799 TCCTGCCCTACTTGCGGTCGAATAGAGATTGATTTAATTCGTATCGCTAATGAAGTGGAA 858

Qy        289 GlnArgLeuGluAspIleIleThrProMetAspValSerIleIleGlyCysValValAsn 308
Db        859 AATTACATAGCAAAGATTGAAGTTCCGATTAAAGTAGCCGTGCTTGGCTGTGCGGTCAAC 918

Qy        309 GlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyGlyAsnLysLysSerGly 328
Db        919 GGCCCTGGAGAAGCTCGCGAAGCCGATATCGGAATTGCTGGTTCAAACGGAGAAGGCCTT 978

Qy        329 LeuTyrGluAspGlyValArgLysAspArgLeuAspAsnAsnAspMetIleAspGlnLeu 348
Db        979 CTTTTTAGACATGGTAAAATTATTCGAAAAGTACCGGAAGCTATTATGATAGACGAACTT 1038

Qy        349 GluAlaArgIleArgAlaLysAlaSerGlnLeuAspGluAlaArgArgIleAspValGln 368
```

```
Db        1039 AAGAAAGAAATTGATATTTTGGCAGAAGAATTT---TTTGTGAAGAAAATAGATTTGGAA 1095

Qy         369 GlnVal 370
Db        1096 AGCCTT 1101
```

RESULT 10
ABA03041
ID ABA03041 standard; DNA; 2944528 BP.
XX
AC ABA03041;
XX
DT 05-FEB-2002 (first entry)
XX
DE Listeria monocytogenes EGD-e genome sequence.
XX
KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
KW vitamin B12; bacterial infection; disease; ds.
XX
OS Listeria monocytogenes.
XX
PN WO200177335-A2.
XX
PD 18-OCT-2001.
XX
PF 11-APR-2001; 2001WO-FR01118.
XX
PR 11-APR-2000; 2000FR-0004629.
XX
PA (INSP) INST PASTEUR.
XX
PI Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;
PI Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;
PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;
PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
PI Maduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;
PI Rose M, Voss H;
XX
DR WPI; 2002-010914/01.
XX
PT Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
PT and prevention of Listeria and related bacterial infections, and
PT related polypeptides -
XX
PS Claim 1; SEQ ID No 1; 192pp; French.
XX
CC The present sequence is the genome sequence of Listeria monocytogenes
CC EGD-e. This sequence and fragments of this sequence are useful for
CC selecting probes and primers for detecting genes in L. monocytogenes and
CC related organisms, and to study genetic polymorphisms and other genomes.
CC Proteins (ABB47297-ABB50149) expressed from the present sequence are
CC useful for raising specific antibodies, identification of L.
CC monocytogenes and related organisms, and for biosynthesis and
CC biodegradation, especially biosynthesis of Vitamin B12. This sequence and
CC proteins encoded by it are also useful for selecting compounds that
CC regulate gene expression and cell replication and modulate L.
CC monocytogenes-related diseases. In addition, this sequence and proteins
CC encoded by it are useful in pharmaceutical and vaccines compositions for
CC the treatment or prevention of infections by L. monocytogenes and related
CC organisms.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2944528 BP; 914202 A; 563301 C; 555061 G; 911964 T; 0 other;

Alignment Scores:

Pred. No.:	9.72e-76	Length:	2944528
Score:	868.00	Matches:	178
Percent Similarity:	67.68%	Conservative:	67
Best Local Similarity:	49.17%	Mismatches:	115

```
Query Match:          46.52%      Indels:        2
DB:                   24          Gaps:          2

US-09-921-992-78 (1-372) x ABA03041 (1-2944528)

Qy         10 ArgLysSerThrArg---IleTyrValGlyAsnValProIleGlyAspGlyAlaProIle 28
              |||:::::|||||   :::   ||||||||||:::   ||||||           :::
Db    1474820 CGCGAAAACACTCGCCCAGTCCAAGTGGGTAATTTAACTATTGGTGGTAGTGAGGAATTA 1474879

Qy         29 AlaValGlnSerMetThrAsnThrArgThrThrAspValGluAlaThrValAsnGlnIle 48
                ::::|||||||||||||   |||:::|||   ||||||||||||||||||   :::|||
Db    1474880 ACTATCCAAAGCATGACTACTACAAAGACACATGATGTCGAAGCAACAGTAGCAGAAATT 1474939

Qy         49 LysAlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMetAspAla 68
                  ||||||      |||      ||||||||||||:::   |||          |||
Db    1474940 CACCGATTAGAAGAAGCTGGTTGTCAGATTGTGCGAGTTGCTTGTCCTGATGAACGTGCA 1474999

Qy         69 AlaGluAlaPheLysLeuIleLysGlnGlnValAsnValProLeuValAlaAspIleHis 88
              |||   |||         |||||:::::::::::::::::||||||||||||||||||||
Db    1475000 GCGAATGCTCTTAGTGCCATCAAGAAAAAGATTCATATTCCGCTTGTGGCAGATATTCAT 1475059

Qy         89 PheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArgIleAsn 108
              ||||||||||||:::|||||||||      :::   |||||||||   :::||||||||||
Db    1475060 TTTGATTACCGACTAGCACTTAAAGCTATTGATGCAGGTGTTGACAAAATTCGGATTAAT 1475119

Qy        109 ProGlyAsnIleGlyAsnGluGluArgIleArgMetValValAspCysAlaArgAspLys 128
              ||||||||||||||||      :::|||:::      ||||||:::   |||:::   |||
Db    1475120 CCTGGTAACATTGGTCGCCGTGATCGGGTGGAAAAAGTGGTTAATGCTGCTAAAGCAAAA 1475179

Qy        129 AsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGlnGluLys 148
              ||||||||||||||||||||||||||||||||||||||||||||||   :::   :::|||
Db    1475180 AATATTCCAATCCGTATTGGGGTTAATGCTGGTAGTTTAGAAAAGAAAATTATTCAAAAA 1475239

Qy        149 TyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeu 168
              ||||||   ||||||         ::::::||||||||:::   |||:::      |||
Db    1475240 TATGGTTACCCTACTGCTGACGGAATGGTAGAAAGTGCACTTGCCCATATTAAAATTCTC 1475299

Qy        169 AspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAla 188
              :::   |||:::|||            ||||||:::|||||||||||||||   ||||||
Db    1475300 GAAGATTTAGATTTTTATGATATCATTGTTTCTTTGAAGGCTTCTGATGTGAATTTAGCA 1475359

Qy        189 ValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeuGlyIleThr 208
              :::|||:::|||         ::::::      :::   ||||||||||||||||||||
Db    1475360 ATTGAAGCTTATGATAAAGCTAGTCGCGCATTTAATTATCCTCTGCATCTCGGAATTACA 1475419

Qy        209 GluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeuLeuLeu 228
              |||:::|||         :::|||   :::||||||||   ||||||||   :::|||
Db    1475420 GAATCTGGTACACAATTTGCTGGAGGAATAAAAAGTGCTGCTGGTTTAGGAGCGATACTC 1475479

Qy        229 SerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGluGluIle 248
              |||   |||||||||:::||||||||||||||||||:::||||||||||||||||||||
Db    1475480 AGTTTGGGCATTGGAAATACATTACGAGTATCTTTGAGTGCTGATCCTGTGGAAGAAATA 1475539

Qy        249 LysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsnPheIle 268
              ||||||      ::::::|||||||||      :::   |||               |||
Db    1475540 AAAGTGGCCCGGGAAGTTTTAAAATCATTTGGTCTTTCCTCGAATGCCGCAATGCTTATC 1475599

Qy        269 AlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsnAlaLeuGlu 288
              :::||||||||||||   |||   |||   |||:::|||         |||   :::|||
Db    1475600 TCCTGCCCTACTTGCGGTCGAATAGAGATTGATTTAATTCGTATCGCTAATGAAGTGGAA 1475659

Qy        289 GlnArgLeuGluAspIleIleThrProMetAspValSerIleIleGlyCysValValAsn 308
                 :::      |||      |||:::   |||:::::::::::||||||   ||||||
Db    1475660 AATTACATAGCAAAGATTGAAGTTCCGATTAAAGTAGCCGTGCTTGGCTGTGCGGTCAAC 1475719

Qy        309 GlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyGlyAsnLysLysSerGly 328
              ||||||||||||||      :::   :::|||:::   |||   |||   :::
Db    1475720 GGCCCTGGAGAAGCTCGCGAAGCCGATATCGGAATTGCTGGTTCAAACGGAGAAGGCCTT 1475779

Qy        329 LeuTyrGluAspGlyValArgLysAspArgLeuAspAsnAsnAspMetIleAspGlnLeu 348
              |||:::      |||            ::::::            ||||||||||:::|||
Db    1475780 CTTTTTAGACATGGTAAAATTATTCGAAAAGTACCGGAAGCTATTATGATAGACGAACTT 1475839

Qy        349 GluAlaArgIleArgAlaLysAlaSerGlnLeuAspGluAlaArgArgIleAspValGln 368
                 :::      |||         |||   :::            :::::||||||::::::
Db    1475840 AAGAAAGAAATTGATATTTTGGCAGAAGAATTT---TTTGTGAAGAAAATAGATTTGGAA 1475896

Qy        369 GlnVal 370
                 :::
Db    1475897 AGCCTT 1475902
```

RESULT 11
ABQ70939
ID ABQ70939 standard; DNA; 6157 BP.
XX
AC ABQ70939;
XX
DT 29-AUG-2002 (first entry)
XX
DE Listeria monocytogenes 4b contig DNA sequence #881.
XX
KW Antibacterial; Listeria; food contamination; mutational analysis;
KW infection; ds.
XX
OS Listeria monocytogenes 4b.
XX
PN WO200228891-A2.
XX
PD 11-APR-2002.
XX
PF 04-OCT-2001; 2001WO-FR03061.
XX
PR 04-OCT-2000; 2000FR-0012697.
XX
PA (INSP) INST PASTEUR.
PA (CNRS) CNRS CENT NAT RECH SCI.
XX
PI Kunst F, Glaser P;
XX
DR WPI; 2002-332479/37.
XX
PT New genomic sequences from Listeria species, useful for detection,
PT treatment and prevention of infection, also related polypeptides,
PT antibodies and modulators -
XX
PS Claim 14; SEQ ID 3752; 180pp; French.
XX
CC The present invention relates to nucleic acid sequences
CC (ABQ67188-ABQ71212) from Listeria sp. The sequences are useful as probes
CC and primers for identification and/or detection of Listeria (e.g. as
CC contaminants in foods, or mutational analysis) and for analysis of
CC gene expression. Proteins encoded by the nucleic acid sequences can be
CC used to screen for compounds that modulate gene expression, replication
CC and pathogenicity of Listeria (potential therapeutic agents), also for
CC treating infections by Listeria, and are useful as immunogens in
CC anti-Listeria vaccines.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 6157 BP; 1731 A; 1065 C; 1209 G; 2150 T; 2 other;

```
Alignment Scores:
Pred. No.:              1.12e-78      Length:          6157
Score:                  861.50        Matches:         175
Percent Similarity:     68.66%        Conservative:    66
Best Local Similarity:  49.86%        Mismatches:      109
Query Match:            46.17%        Indels:          1
DB:                     24            Gaps:            1

US-09-921-992-78 (1-372) x ABQ70939 (1-6157)

Qy         10 ArgLysSerThrArg---IleTyrValGlyAsnValProIleGlyAspGlyAlaProIle 28
              |||:::::|||||   :::   ||||||||||:::   ||||||           :::
Db       3123 CGCGAAAACACTCGGCCAGTCCAAGTAGGTAATTTAACTATTGGTGGTAGTGAGGAATTA 3182
```

```
Qy        29 AlaValGlnSerMetThrAsnThrArgThrThrAspValGluAlaThrValAsnGlnIle 48
               :::|||||||||||||   |||:::|||   |||||||||||||||||    :::|||
Db      3183 ACTATCCAAAGCATGACTACTACAAAGACACATGATGTCGAAGCAACAGTAGCGGAAATT 3242

Qy        49 LysAlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMetAspAla 68
                   ||||||   ||||||      |||||||||||:::    |||        |||
Db      3243 CACCGATTAGAAGAGGTGGGTTGTCAGATTGTGCGAGTTGCTTGTCCTGATGAGCGTGCA 3302

Qy        69 AlaGluAlaPheLysLeuIleLysGlnGlnValAsnValProLeuValAlaAspIleHis 88
             |||   |||      |||||||::::::::::|||||||||||||||||||||||||
Db      3303 GCCAATGCTCTTAGTGCCATCAAGAAAAGGATTCATATTCCTCTTGTGGCAGATATTCAT 3362

Qy        89 PheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArgIleAsn 108
             ||||||||||||:::|||||||||      :::   ||||||||    :::||||||||||
Db      3363 TTTGATTACCGACTAGCACTTAAAGCTATTGATGCAGGTGTTGACAAAATCCGGATTAAT 3422

Qy       109 ProGlyAsnIleGlyAsnGluGluArgIleArgMetValValAspCysAlaArgAspLys 128
             |||||||||||||||       :::|||:::      |||||||:::   |||:::   |||
Db      3423 CCTGGTAACATCGGTCGCCGTGATCGGGTGGAAAAAGTGGTTAATGCTGCTAAAGCAAAA 3482

Qy       129 AsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGlnGluLys 148
             ||||||||||||||||||||||||||||||||||||||||||||||   :::    :::|||
Db      3483 AATATTCCAATCCGTATTGGGGTTAATGCTGGTAGTTTAGAAAAGAAAATTATTCAAAAA 3542

Qy       149 TyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeu 168
             ||||||     ||||||   :::   ::::::|||||||||:::    |||:::       |||
Db      3543 TACGGTTACCCTACTGCTGAAGGAATGGTAGAAAGTGCACTTGATCATATTAAAATTCTC 3602

Qy       169 AspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAla 188
             :::    |||:::|||            ||||||:::||||||||||||||||    ||||||
Db      3603 GAAGATTTAGATTTTTATGATATCATTGTTTCTTTGAAAGCTTCTGATGTGAATTTAGCA 3662

Qy       189 ValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeuGlyIleThr 208
             :::|||:::|||         ::::::    :::    ||||||||||||||||||||||
Db      3663 ATTGAAGCTTATGATAAAGCTAGTCGCGCATTTAATTATCCTCTACATCTTGGAATTACA 3722

Qy       209 GluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeuLeuLeu 228
             |||:::|||         :::|||   :::|||||||||    ||||||||    :::|||
Db      3723 GAATCTGGTACGCAGTTTGCTGGAGGAATAAAAAGTGCTGCTGGTTTAGGAGCGATACTC 3782

Qy       229 SerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGluGluIle 248
             |||    ||||||||||||||||||||||||||||:::||||||||||||||||||||||
Db      3783 AGTTTGGGCATTGGAAATACATTACGGGTATCTTTGAGTGCTGATCCTGTGGAAGAGATA 3842

Qy       249 LysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsnPheIle 268
             ||||||         :::|||||||||||      :::   |||               |||
Db      3843 AAAGTGGCCCGGGAGGTTTTAAAATCGTTTGGCCTTTCCTCGAATGCCGCCATGCTTATC 3902

Qy       269 AlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsnAlaLeuGlu 288
             :::||||||||||||||   |||    |||   |||:::|||      |||    :::|||
Db      3903 TCCTGCCCTACTTGCGGTCGAATAGAGATTGATTTGATTCGTATCGCTAATGAAGTGGAA 3962

Qy       289 GlnArgLeuGluAspIleIleThrProMetAspValSerIleIleGlyCysValValAsn 308
                   :::      |||      |||:::   |||:::::::::||||||    ||||||
Db      3963 AATTACATAGCAACGATTAAAGCTCCGATTAAAGTAGCCGTGCTTGGCTGTGCGGTCAAC 4022

Qy       309 GlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyGlyAsnLysLysSerGly 328
             |||||||||||||||         :::   :::|||:::    |||    |||     :::
Db      4023 GGTCCCGGAGAAGCTCGCGAAGCAGATATCGGAATTGCTGGTTCAAATGGAGAAGGCCTT 4082

Qy       329 LeuTyrGluAspGlyValArgLysAspArgLeuAspAsnAsnAspMetIleAspGlnLeu 348
             |||:::       |||               ::::::          |||:::::::::|||
Db      4083 CTTTTTAGACATGGTAAAATCATCCGAAAAGTACCAGAAGCCATTATGGTAGAAGAACTT 4142

Qy       349 GluAlaArgIleArgAlaLysAlaSerGlnLeu 359
             :::      |||         |||   ::::::
Db      4143 AAGAAAGGAATTGATATTTTGGCCAGAAGAATA 4175
```

RESULT 12
ABK74401
ID   ABK74401 standard; DNA; 1083 BP.
XX
AC   ABK74401;
XX
DT   13-AUG-2002 (first entry)
XX
DE   Bacillus licheniformis genomic sequence tag (GST) #1692.
XX
KW   Differential gene expression; genomic sequenced tag; GST;
KW   altered culture condition; environmental stress;
KW   physiological provocation; ds.
XX
OS   Bacillus licheniformis.
XX
PN   WO200229113-A2.
XX
PD   11-APR-2002.
XX
PF   05-OCT-2001; 2001WO-US31437.
XX
PR   06-OCT-2000; 2000US-0680598.
PR   27-MAR-2001; 2001US-279526P.
XX
PA   (NOVO) NOVOZYMES BIOTECH INC.
PA   (NOVO) NOVOZYMES AS.
XX
PI   Berka R, Clausen IG;
XX
DR   WPI; 2002-416684/44.
XX
PT   Monitoring differential expression of several genes in first Bacillus
PT   cell relative to expression of same genes in one or more second
PT   Bacillus cells, by using substrate containing Bacillus genomic
PT   sequenced tag array -
XX
PS   Claim 4; SEQ ID NO 1692; 200pp; English.
XX
CC   The invention describes a method of monitoring differential expression of
CC   genes in a first Bacillus cell relative to expression of the genes in
CC   other Bacillus cells, comprising hybridising labelled nucleic acid probes
CC   isolated from Bacillus cells to a substrate containing array of Bacillus
CC   genomic sequenced tags (GST), examining the array, and determining
CC   relative gene expression by an observed hybridisation reporter signal of
CC   a spot in the array. The method is useful for measuring the expression of
CC   genes in a first Bacillus cell relative to expression of the same genes
CC   in one or more second Bacillus cells. The method is useful for monitoring
CC   global expression of several genes from a Bacillus cell, discovering new
CC   genes, identifying possible functions of unknown open reading frames and
CC   monitoring gene copy number variation and stability. Monitoring changes
CC   in expression of genes may be used to provide a representation of the way
CC   in which Bacillus cells adapt to changes in culture conditions,
CC   environmental stress or other physiological provocation. Extensive
CC   follow-up characterisation is unnecessary, when one spot on an array
CC   equals one gene or one open reading frame, since sequence information is
CC   available. This sequence represents a genomic sequence tag (GST) used in
CC   the method of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1083 BP; 314 A; 278 C; 265 G; 226 T; 0 other;

Alignment Scores:			
Pred. No.:	1.34e-75	Length:	1083
Score:	822.00	Matches:	162
Percent Similarity:	68.45%	Conservative:	55
Best Local Similarity:	51.10%	Mismatches:	100
Query Match:	44.05%	Indels:	0
DB:	24	Gaps:	0

US-09-921-992-78 (1-372) x ABK74401 (1-1083)

```
Qy         9 ArgArgLysSerThrArgIleTyrValGlyAsnValProIleGlyAspGlyAlaProIle 28
                |||    |||:::       :::   ||||||    :::    ||||||         :::
Db        24 CGTACAAAAACGCGTCCCGTTAAAGTGGGACCTTTAACAATAGGCGGCAATAACGAAGTC 83
```

```
Qy         29 AlaValGlnSerMetThrAsnThrArgThrThrAspValGluAlaThrValAsnGlnIle 48
              :::|||||||||||||   |||:::|||   |||||||||||||||||   :::|||
Db         84 GTCATTCAAAGCATGACAACAACGAAAACACATGACGTTGAAGCAACCGTCGCCGAAATC 143

Qy         49 LysAlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMetAspAla 68
                 |||         |||       |||||||||||:::   |||         |||
Db        144 AACAGACTCGCGGAAGCAGGATGTCAAATCGTCCGCGTCGCCTGTCCTGATGAACGGGCT 203

Qy         69 AlaGluAlaPheLysLeuIleLysGlnGlnValAsnValProLeuValAlaAspIleHis 88
              |||:::|||        |||||||||||||||||||||||||   |||||||||
Db        204 GCCGACGCCATTCCAGAGATCAAAAAGCGGATATCCATCCCTCTTGTCGTGGATATTCAT 263

Qy         89 PheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArgIleAsn 108
              |||:::|||||||:|||||||||       |||   |||   |||   :::|||||||||
Db        264 TTCAACTATAAATTGGCATTAAAAGCGATCGAAGGCGGAGCCGATAAAATCCGCATCAAT 323

Qy        109 ProGlyAsnIleGlyAsnGluGluArgIleArgMetValValAspCysAlaArgAspLys 128
              ||||||||||||||       |||:::::       |||||||:::   |||:::::::|||
Db        324 CCGGGTAACATCGGCCGCCGCGAAAAGGTTGAAGCGGTCGTCAACGCAGCGAAGGAAAAG 383

Qy        129 AsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGlnGluLys 148
              ||||||||||||||||||||||||||||||||||||||||||||||   :::   ||||||
Db        384 GGCATTCCGATCCGGATCGGCGTCAATGCAGGCTCTCTGGAAAAACGAATCCTTGAGAAG 443

Qy        149 TyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeu 168
              ||||||   ||||||         ::::::|||||||||:::   |||:::      |||
Db        444 TACGGCTATCCGACAGCAGACGGCATGGTCGAAAGCGCCCTGCACCACATTAAAATTCTT 503

Qy        169 AspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAla 188
              :::   |||:::|||            ||||||:::|||||||||||||||   ||||||
Db        504 GAGGATCTCGATTTCCACGATATCATCGTCAGCATGAAGGCGTCTGATGTAAACCTGGCG 563

Qy        189 ValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeuGlyIleThr 208
              :::|||:::|||        ||||||      |||   ||||||||||||||||||||
Db        564 ATTGAGGCATATGAAAAAGCGGCTAAAGCTTTCGATTATCCGCTTCATTTAGGCATCACC 623

Qy        209 GluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeuLeuLeu 228
              |||:::|||          :::|||   ||||||||||||   |||||||||   :::|||
Db        624 GAATCGGGAACGCTGTTTGCCGGTACAGTGAAAAGCGCCGCGGGTCTCGGCGCGATCCTT 683

Qy        229 SerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGluGluIle 248
              |||:::|||||||||:::||||||||||:::|||||||:::|||||||||||||||||:::
Db        684 TCAAAAGGAATCGGCAATACATTGCGGATTTCTTTAAGCGCGGACCCGGTCGAAGAAGTA 743

Qy        249 LysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsnPheIle 268
              ||||||      ::::::|||||||||      :::   |||                |||
Db        744 AAAGTCGCCAGGGAGCTCCTGAAATCTTTCGGGCTCGCTTCCAATGCGGCAACATTGATT 803

Qy        269 AlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsnAlaLeuGlu 288
              :::||||||||||||   |||   |||   |||:::|||          |||   |||
Db        804 TCCTGCCCGACTTGCGGCCGGATCGAAATCGATTTGATCTGCATTGCCAATGAATTCGAA 863

Qy        289 GlnArgLeuGluAspIleIleThrProMetAspValSerIleIleGlyCysValValAsn 308
                  :::      |||      |||:::  |||::::::::::::||||||   ||||||
Db        864 GATTACATCGCGAAAATCAAGGCGCCGATCAAAGTTGCCGTTCTCGGCTGTGCCGTCAAC 923

Qy        309 GlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyGlyAsnLys 325
              ||||||||||||||      :::   :::   :::   |||      :::
Db        924 GGTCCCGGAGAAGCGCGCGAAGCCGATATCCGGATCGCCGGCGCACGCCGG 974

RESULT 13
ABQ81842
ID   ABQ81842 standard; DNA; 349980 BP.
XX
AC   ABQ81842;
XX
DT   19-NOV-2002  (first entry)
XX
DE   Bifidobacterium longum NCC2705 genomic sequence SEQ ID NO:1.
XX
KW   Bifidobacterium longum NCC2705; Bifidobacterium; bacterial;
KW   antidiarrheic; antibacterial; inhibitor of Salmonella; detection;
KW   identification; lactic acid bacterium; diarrhoea; pathogenic bacteria;
KW   rotavirus; food composition; pharmaceutical composition; gene; ds.
XX
OS   Bifidobacterium longum.
XX
PN   EP1227152-A1.
XX
PD   31-JUL-2002.
XX
PF   30-JAN-2001; 2001EP-0102050.
XX
PR   30-JAN-2001; 2001EP-0102050.
XX
PA   (NEST ) SOC PROD NESTLE SA.
XX
DR   WPI; 2002-668397/72.
XX
PT   Novel polynucleotide comprising Bifidobacterium genome sequence useful
PT   as a probe or primer for detecting and/or identifying Bifidobacterium
PT   longum in a biological sample  -
XX
PS   Claim 1; SEQ ID 1; 80pp; English.
XX
CC   The present invention describes a polynucleotide (I) comprising a
CC   sequence of a Bifidobacterium genome selected from the nucleotide
CC   sequences given in ABQ81842 and ABQ81843, or a sequence exhibiting at
CC   least 90% identity or which hybridises with the sequences given in
CC   ABQ81842 and ABQ81843. Also described is a polynucleotide (II) encoding
CC   a fusion protein, comprising a sequence selected from 1097 sequences
CC   given in ABP65258 to ABP66354 ligated in frame to a polynucleotide
CC   encoding a heterologous polypeptide. (I) has antidiarrheic and
CC   antibacterial activities, and can be used as an inhibitor of Salmonella.
CC   (I) (which is a probe) is useful for the detection and/or identification
CC   of Bifidobacterium longum in a biological sample. A carrier containing
CC   the lactic acid bacterium Bifidobacterium longum NCC2705 (CNCM I-2618)
CC   can be used for preventing and/or treating diarrhoea brought about by
CC   pathogenic bacteria and/or rotavirus. The carrier is a food composition
CC   selected from milk, yogurt, curd, cheese, fermented milks, milk based
CC   fermented products, ice-creams, fermented cereal based products, milk
CC   based powders, infant formula, pet food or a pharmaceutical composition
CC   selected from tablets, liquid bacterial suspensions, dried oral
CC   supplement, wet oral supplement, dry tube feeding or wet tube feeding.
CC   (I) is useful in DNA arrays or chips to carry out analysis of the
CC   expression of the Bifidobacterium gene. ABQ81844 to ABQ81850 represent
CC   Bifidobacterium related nucleotide sequences given in the Sequence
CC   Listing from the present invention but not mentioned further within the
CC   specification.
CC   N.B. The sequence data for this patent is not represented in the printed
CC   specification but is based on sequence information supplied by the
CC   European Patent Office.
XX
SQ   Sequence 349980 BP; 72540 A; 102738 C; 103221 G; 71481 T; 0 other;

Alignment Scores:
Pred. No.:              1.95e-71        Length:         349980
Score:                  814.50          Matches:        167
Percent Similarity:     65.65%          Conservative:   70
Best Local Similarity:  46.26%          Mismatches:     111
Query Match:            43.65%          Indels:         13
DB:                     24              Gaps:           4

US-09-921-992-78 (1-372) x ABQ81842 (1-349980)

Qy          3 AsnGlnAlaProIleGln---ArgArgLysSerThrArgIleTyrValGlyAsnValPro 21
              ::::::::::|||:::      ||||||||||||   ||||||   ||||||   ||||||
Db     118761 AGCGAATCGCCCTTGCACCCGCGCCGCAAGTCCCGCCGCATCATGGTGGGTCCGGTGCCG 118820

Qy         22 IleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrThrAspVal 41
              :::|||   ||||||||||||:::|||||||||||||||||||||   |||   :::|||
Db     118821 GTGGGCGGGGGAGCGCCCATCTCCGTGCAGTCGATGACCAACACGCTCACGGCCAACGTA 118880

Qy         42 GluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArgVal 61
                   |||||||::   ||||||      |||         |||   ||||||||||||||||
```

```
Db     118881 CCTGCCACGTTGCAGCAGATCGCCGAACTGACCGCCGCCGGCTGCGACATCGTGCGCGTG 118940

Qy         62 SerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsnVal 81

Db     118941 GCCGTGCCGAGTCAGGACGATGCCGACGCACTGCCGGAAATCTGCCGCAAATCGCCGATT 119000

Qy         82 ProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyrGly 101

Db     119001 CCGGTCATCGCCGATATCCACTTCCAGTCCAAATACGTGTTCCAGGCCATCGACGCCGGC 119060

Qy        102 ValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMet--- 120

Db     119061 TGCGCGGCCGTGCGCGTGAACCCCGGCAACATCCGTAAGTTCGACGAAGTCGGCCCGGAC 119120

Qy        121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140

Db     119121 ATCTGCAAGGCCGCCACCGACGCAGGCATCTCGCTGCGTATCGGTGTGAACGCCGGCTCG 119180

Qy        141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSer 160

Db     119181 CTGGACAAGGAACTCTACGCCAAATACGGTGGCCCGACTCCGGAAGCGCTGGTGGCCTCC 119240

Qy        161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180

Db     119241 GCCTTGAAGGAAGCGCACATGTTCGAGGACGTCGGCTTCCACGATTTCAAGATCTCCGTC 119300

Qy        181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200

Db     119301 AAACACCATGACGTCATCACCATGGTCGAAACGTACCGACTGCTCGCTTCCAAGGGCGAT 119360

Qy        201 GlnProLeuHisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSer 220

Db     119361 TGGCCGCTGCACCTCGGCGTCACCGAGGCCGGCCCTGCTTGGCAAGGCACCATCAAATCC 119420

Qy        221 AlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240

Db     119421 TGCCTGGCGTTCGGTGCGCTCTTGGCCGAGGGCATTGGCGACACGATTCGCGTGTCTCTC 119480

Qy        241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260

Db     119481 TCCGCGCCGCCGGCCGAAGAGGTCAAGGTGGGTTGCAAACTCTTGGAATACATGGGATTG 119540

Qy        261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspVal 280

Db     119541 CGCCCCCGCAAGTTCGACATCATCTCCTGCCCGAGCTGCGGCCGCGCCCAAGTGGATGTG 119600

Qy        281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300

Db     119601 ATCCAGCTCGCCTCCGCCGTCACCGAAGGCCTGAAGGACGTGACCGCGCCGATCCGAGTG 119660

Qy        301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320

Db     119661 GCCGTCATGGGCTGCATCGTCAACGGTCCGGGCGAGGCACGCGAAGCCGACCTCGGCGTG 119720

Qy        321 ThrGlyGlyAsnLysLysSerGlyLeuTyrGluAspGly----------ValArg 335

Db     119721 GCCTCCGGCAATGGCAAGGGCCAGATCTTCATCAAGGGCAAGGTCATCAAGACCGTGCCC 119780

Qy        336 LysAspArgLeuAspAsn----------AsnAspMetIleAspGlnLeuGlu 349

Db     119781 GAAGACCAGATCGTCGACACGCTGCTCACCATAGCGAACGATATCGCCGCCCAAATGGAG 119840

Qy        350 Ala 350

Db     119841 GCC 119843

RESULT 14
ABQ68324
ID   ABQ68324 standard; DNA; 994 BP.
XX
AC   ABQ68324;
XX
DT   29-AUG-2002  (first entry)
XX
DE   Listeria monocytogenes 4b contig DNA sequence #1090.
XX
KW   Antibacterial; Listeria; food contamination; mutational analysis;
KW   infection; ds.
XX
OS   Listeria monocytogenes 4b.
XX
PN   WO200228891-A2.
XX
PD   11-APR-2002.
XX
PF   04-OCT-2001; 2001WO-FR03061.
XX
PR   04-OCT-2000; 2000FR-0012697.
XX
PA   (INSP ) INST PASTEUR.
PA   (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI   Kunst F,  Glaser P;
XX
DR   WPI; 2002-332479/37.
XX
PT   New genomic sequences from Listeria species, useful for detection,
PT   treatment and prevention of infection, also related polypeptides,
PT   antibodies and modulators -
XX
PS   Claim 14; SEQ ID 1137; 180pp; French.
XX
CC   The present invention relates to nucleic acid sequences
CC   (ABQ67188-ABQ71212) from Listeria sp. The sequences are useful as probes
CC   and primers for identification and/or detection of Listeria (e.g. as
CC   contaminants in foods, or mutational analysis) and for analysis of
CC   gene expression. Proteins encoded by the nucleic acid sequences can be
CC   used to screen for compounds that modulate gene expression, replication
CC   and pathogenicity of Listeria (potential therapeutic agents), also for
CC   treating infections by Listeria, and are useful as immunogens in
CC   anti-Listeria vaccines.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic format
CC   directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 994 BP; 296 A; 169 C; 239 G; 290 T; 0 other;

Alignment Scores:
Pred. No.:                1.3e-74        Length:          994
Score:                    812.00         Matches:         164
Percent Similarity:       70.32%         Conservative:    54
Best Local Similarity:    52.90%         Mismatches:      90
Query Match:              43.52%         Indels:          2
DB:                       24             Gaps:            1

US-09-921-992-78 (1-372) x ABQ68324 (1-994)

Qy         29 AlaValGlnSerMetThrAsnThrArgThrThrAspValGluAlaThrValAsnGlnIle 48

Db          4 GCTGGAAAGAGCATGACTACTACAAAGCTACATGATGTCGAAGCAACAGTAGCGGAAATT 63

Qy         49 LysAlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMetAspAla 68

Db         64 CACCGATTAGAAGAGGTGGGTTGTCAGATTGTGCGAGTTGCTTGTCCTGATGAGCGTGCA 123

Qy         69 AlaGluAlaPheLysLeuIleLysGlnGlnValAsnValProLeuValAlaAspIleHis 88

Db        124 GCCAATGCTCTTAGTGCCATCAAGAAAAGGATTCATATTCCTCTTGTGGCAGATATTCAT 183

Qy         89 PheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArgIleAsn 108

Db        184 TTTGATTACCGACTAGCACTTAAAGCTATTGATGCAGGTGTTGACAAAATCCGGATTAAT 243

Qy        109 ProGlyAsnIleGlyAsnGluGluArgIleArgMetValValAspCysAlaArgAspLys 128

Db        244 CCTGGTAACATCGGTCGCCGTGATCGGGTGGAAAAAGTGGTTAATGCTGCTAAAGCAAAA 303
```

```
Qy        129 AsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGlnGluLys 148
              ||||||||||||||||||||||||||||||||||||||||||||||||   :::   :::|||
Db        304 AATATTCCAATCCGTATTGGGGTTAATGCTGGTAGTTTAGAAAAGAAAATTATTCAAAAA 363

Qy        149 TyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeu 168
              ||||||   ||||||  :::   ::::::||||||||||:::   |||:::      |||
Db        364 TACGGTTACCCTACTGCTGAAGGAATGGTAGAAAGTGCACTTGATCATATTAAAATTCTC 423

Qy        169 AspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAla 188
              :::   |||:::|||           ||||||:::|||||||||||||||   ||||||
Db        424 GAAGATTTAGATTTTTATGATATCATTGTTTCTTTGAAAGCTTCTGATGTGAATTTAGCA 483

Qy        189 ValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeuGlyIleThr 208
              :::|||:::|||         ::::::      :::   ||||||||||||||||||||
Db        484 ATTGAAGCTTATGATAAAGCTAGTCGCGCATTTAATTATCCTCTACATCTTGGAATTACA 543

Qy        209 GluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeuLeuLeu 228
              |||:::|||         :::|||   :::|||||||||   ||||||||   :::|||
Db        544 GAATCTGGTACGCAGTTTGCTGGAGGAATAAAAAGTGCTGCTGGTTTAGGAGCGATACTC 603

Qy        229 SerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGluGluIle 248
              |||   |||||||||:::||||||||||||||||||:::||||||||||||||||||||
Db        604 AGTTTGGGCATTGGAAATACATTACGGGTATCTTTGAGTGCTGATCCTGTGGAAGAGATA 663

Qy        249 LysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsnPheIle 268
              ||||||      ::::::|||||||||      :::   |||                |||
Db        664 AAAGTGGCCCGGGAGGTTTTAAAATCGTTTGGCCTTTCCTCGAATGCCGCCATGCTTATC 723

Qy        269 AlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsnAlaLeuGlu 288
              :::||||||||||||   |||   |||   |||:::|||           |||   :::|||
Db        724 TCCTGCCCTACTTGCGGTCGAATAGAGATTGATTTGATTCGTATCGCTAATGAAGTGGAA 783

Qy        289 GlnArgLeuGluAspIleIleThrProMetAspValSerIleIleGlyCysValValAsn 308
                  :::      |||      |||:::   |||:::::::::::||||||   ||||||
Db        784 AATTACATAGCAACGATTAAAGCTCCGATTAAAGTAGCCGTGCTTGGCTGTGCGGTCAAC 843

Qy        309 GlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyGlyAsnLysLysSerGly 328
              ||||||||||||||      :::   :::|||:::   |||   |||   :::
Db        844 GGTCCCGGAGAAGCTCGCGAAGCAGATATCGGAATTGCTGGTTCAAATGGAGAAGGCCTT 903

Qy        329 LeuTyrGluAspGly------ValArgLys 336
              |||:::     |||     :::|||||||
Db        904 CTTTTTAGACATGGTAAAATCATCCGAAAA 933

RESULT 15
AAX14017
ID   AAX14017 standard; DNA; 1180 BP.
XX
AC   AAX14017;
XX
DT   31-MAR-1999  (first entry)
XX
DE   H. pylori GHPO 76 gene.
XX
KW   GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
KW   peptic ulcer disease; ss.
XX
OS   Helicobacter pylori.
XX
FH   Key             Location/Qualifiers
FT   CDS             51..1130
FT                   /*tag= a
XX
PN   WO9843478-A1.
XX
PD   08-OCT-1998.
XX
PF   01-APR-1998;    98WO-US06371.
XX
PR   29-JUL-1997;    97US-0902615.
PR   01-APR-1997;    97US-0833457.
PR   24-JUN-1997;    97US-0881227.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (INMR ) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
XX
PI   Al-Garawi A,  Kleanthous H,  Miller C,  Oomen RP,  Tomb J;
XX
DR   WPI; 1998-542293/46.
DR   P-PSDB; AAW98298.
XX
PT   New isolated Helicobacter polynucleotides - used to develop products
PT   for the diagnosis, prevention and treatment of Helicobacter
PT   infections and gastrointestinal diseases
XX
PS   Claim 1; Page 372-374; 2054pp; English.
XX
CC   This sequence represents a polynucleotide of the invention. It was
CC   isolated from Helicobacter pylori and encodes a H.pylori GHPO protein.
CC   The polypeptides can be used for preventing or treating Helicobacter
CC   infections, and gastroduodenal diseases associated with these
CC   infections, including acute, chronic, and atrophic gastritis, and peptic
CC   ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used
CC   for the production of antibodies. The products can also be used for
CC   detection and diagnosis.
XX
SQ   Sequence 1180 BP; 380 A; 206 C; 289 G; 305 T; 0 other;

Alignment Scores:
Pred. No.:                1.21e-72      Length:          1180
Score:                    794.00        Matches:         162
Percent Similarity:       66.38%        Conservative:    73
Best Local Similarity:    45.76%        Mismatches:      117
Query Match:              42.55%        Indels:          2
DB:                       19            Gaps:            2

US-09-921-992-78 (1-372) x AAX14017 (1-1180)

Qy          9 ArgArgLysSerThrArgIleTyrValGlyAsnValProIleGlyAspGlyAlaProIle 28
              |||   |||:::   :::|||:::::::|||   |||   ||||||      |||||||||
Db         63 AGAGTTAAGACCAAGCAAATTTTTATCGGTGGCGTGGCCATAGGGGGTGATGCTCCCATA 122

Qy         29 AlaValGlnSerMetThrAsnThrArgThrThrAspValGluAlaThrValAsnGlnIle 48
              :::   ||||||||||||  :::::::|||   |||:::|||:::|||   |||||||||
Db        123 AGCACGCAAAGCATGACCTTTAGCAAAACCGCTGATATTGAAAGCACTAAAAATCAAATT 182

Qy         49 LysAlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMetAspAla 68
                 |||:::     |||||||||:::|||||||||:::|||
Db        183 GACAGACTCAAACTCGCCGGGGCCGATTTAGTGAGGGTGGCGGTGAGTAATGAAAAGGAC 242

Qy         69 AlaGluAlaPheLysLeuIleLysGlnGlnValAsnValProLeuValAlaAspIleHis 88
              |||   |||   |||   :::|||:::          :::|||||||:::||||||||||||
Db        243 GCTCTAGCCTTAAAAGAATTGAAAAAAGTGTCCCCTTTGCCTTTAATCGCTGATATTCAT 302

Qy         89 PheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArgIleAsn 108
              |||   |||:::   ||||||  .:::|||          ||||||   :::||||||||||
Db        303 TTCCATTATAAATTCGCTCTC---ATTGCCGCTCAAAGCGTGGATGCGATCAGGATTAAC 359

Qy        109 ProGlyAsnIleGlyAsnGluGluArgIleArgMetValValAspCysAlaArgAspLys 128
              ||||||||||||||:::::|||:::|||:::   |||||||||      :::::|||
Db        360 CCCGGAAACATCGGCTCTAAAGAGAAGATCAAAGCGGTGGTTGATGCTTGTAAAGAAAAA 419

Qy        129 AsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGlnGluLys 148
              ||||||||||||||||||||||||||||||||||||||||||||||||||||   :::|||
Db        420 AACATTCCTATAAGAATTGGCGTGAATGCTGGGAGTTTAGAAAAGCAGTTTGATCAAAAA 479

Qy        149 TyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeu 168
              ||||||   |||||||||:::   ::::::||||||||||:::   :::         |||
Db        480 TACGGA---CCCACCCCAAAAGGCATGGTAGAAAGCGCTTTGTATAACGCCAAACTTTTA 536

Qy        169 AspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAla 188
              :::   |||:::|||       ||||||:::|||:::|||||||||||||||||
Db        537 GAAGATTTGGATTTTACCAATTTTAAGATTTCTTTAAAAGCGAGCGATGTGATTCGCACC 596

Qy        189 ValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeuGlyIleThr 208
```

```
Db        597 ATAGAAGCTTACAGGATGCTTCGCCCTCTTGTGATCTATCCTTTCCATTTGGGGGTTACG 656

Qy        209 GluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeuLeuLeu 228

Db        657 GAGGCGGGGAATCTTTTTAGCTCCAGTATCAAATCCGCTATGGCTTTAGGGGGGCTTTTA 716

Qy        229 SerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGluGluIle 248

Db        717 ATGGAGGGCATTGGGGATACGATGCGCGTATCCATCACAGGGGAATTAGAAAATGAAATC 776

Qy        249 LysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsnPheIle 268

Db        777 AAAGTGGCCAGAGCAATTTTACGCCATAGCGGGCGGTTGAAAGAAGGGATTAATTGGATT 836

Qy        269 AlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsnAlaLeuGlu 288

Db        837 TCTTGCCCCACTTGCGGGCGCATTGAAGCCAATTTAGTGGATATGGCGATCAAGGTAGAA 896

Qy        289 GlnArgLeuGluAspIleIleThrProMetAspValSerIleIleGlyCysValValAsn 308

Db        897 AAACGCTTAAGCCACATCAAAACCCCTTTAGACATTAGCGTGATGGGTTGCGTGGTGAAT 956

Qy        309 GlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyGlyAsnLysLysSerGly 328

Db        957 GCTTTGGGTGAAGCCAAGCATGCAGACATGGCGATCGCTTTTGGGAATCGCAGCGGTTTG 1016

Qy        329 LeuTyrGluAspGlyValArgLysAspArgLeuAspAsnAsnAspMetIleAspGlnLeu 348

Db       1017 ATCATTAAAGAGGGTAAAGTCATTCACAAACTGGCTGAAAAGGATTTATTTGAAACTTTT 1076

Qy        349 GluAlaArgIleArgAlaLysAlaSerGlnLeuAspGluAla 362

Db       1077 GTGATAGAAGTGGAAAATTTAGCTAAAGAAAGAGAAAAAAGT 1118
```

Search completed: August 15, 2003, 01:38:42
Job time : 3223 secs

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 14, 2003, 23:20:23 ; Search time 83 Seconds
(without alignments)
1978.247 Million cell updates/sec

Title: US-09-921-992-78
Perfect score: 1866
Sequence: 1 MHNQAPIQRRKSTRIYVGNV.....RAKASQLDEARRIDVQQVEK 372

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09921992/runat_12082003_094913_17389/app_query.fasta_1.519
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09921992_@CGN_1_1_56_@runat_12082003_094913_17389 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1691	90.6	1010	2	US-08-827-190-8	Sequence 8, Appli
	2	1691	90.6	1010	4	US-09-170-187-8	Sequence 8, Appli
	3	1624.5	87.1	1345	2	US-08-827-190-4	Sequence 4, Appli
	4	1624.5	87.1	1345	4	US-09-170-187-4	Sequence 4, Appli
	5	1603.5	85.9	1830121	4	US-09-557-884-1	Sequence 1, Appli
	6	1603.5	85.9	1830121	4	US-09-643-990A-1	Sequence 1, Appli
	7	1593.5	85.4	1079	2	US-08-827-190-7	Sequence 7, Appli
	8	1593.5	85.4	1079	4	US-09-170-187-7	Sequence 7, Appli
	9	1372.5	73.6	1644	4	US-09-252-991A-6420	Sequence 6420, Ap
c	10	1372.5	73.6	2202	4	US-09-252-991A-6131	Sequence 6131, Ap
	11	1363.5	73.1	2088	4	US-09-252-991A-6339	Sequence 6339, Ap
c	12	1353	72.5	886	2	US-08-827-190-10	Sequence 10, Appl
c	13	1353	72.5	886	4	US-09-170-187-10	Sequence 10, Appl
	14	1195	64.0	1137	4	US-09-328-352-3780	Sequence 3780, Ap
c	15	812	43.5	886	2	US-08-827-190-9	Sequence 9, Appli
c	16	812	43.5	886	4	US-09-170-187-9	Sequence 9, Appli
	17	764	40.9	555	4	US-09-252-991A-6251	Sequence 6251, Ap
c	18	751.5	40.3	4403765	3	US-09-103-840A-2	Sequence 2, Appli
c	19	751.5	40.3	4411529	3	US-09-103-840A-1	Sequence 1, Appli
c	20	607.5	32.6	543	4	US-09-252-991A-6209	Sequence 6209, Ap
	21	470.5	25.2	1230025	4	US-09-198-452A-1	Sequence 1, Appli
	22	288.5	15.5	1083	4	US-09-221-017B-309	Sequence 309, App
	23	112.5	6.0	3177	4	US-09-252-991A-12534	Sequence 12534, A
	24	111.5	6.0	3998	3	US-09-066-046-5	Sequence 5, Appli
c	25	111	5.9	1665	4	US-09-221-017B-1017	Sequence 1017, Ap
c	26	106.5	5.7	1854	4	US-09-252-991A-14350	Sequence 14350, A
	27	106.5	5.7	2943	4	US-09-252-991A-14267	Sequence 14267, A
c	28	106	5.7	2259	4	US-09-252-991A-784	Sequence 784, App
c	29	106	5.7	1664976	4	US-08-916-421B-1	Sequence 1, Appli
c	30	105.5	5.7	2129	3	US-08-975-762-39	Sequence 39, Appl
c	31	105.5	5.7	2129	3	US-09-295-028-39	Sequence 39, Appl
c	32	105.5	5.7	2129	4	US-09-106-582-39	Sequence 39, Appl
c	33	103.5	5.5	2214	1	US-07-985-458-1	Sequence 1, Appli
	34	101.5	5.4	1694	3	US-09-136-421-11	Sequence 11, Appl
	35	101.5	5.4	2989	1	US-07-927-851-1	Sequence 1, Appli
	36	101.5	5.4	2989	1	US-08-453-323-1	Sequence 1, Appli
	37	101.5	5.4	2989	1	US-08-440-520-1	Sequence 1, Appli
c	38	101	5.4	1149	4	US-09-252-991A-13125	Sequence 13125, A
	39	101	5.4	1398	4	US-09-252-991A-12816	Sequence 12816, A
	40	101	5.4	1572	4	US-09-612-964-1	Sequence 1, Appli
	41	100.5	5.4	1863	4	US-09-328-352-3604	Sequence 3604, Ap
	42	100.5	5.4	2010	3	US-09-039-859-1	Sequence 1, Appli
c	43	100.5	5.4	11601	2	US-08-222-617A-3	Sequence 3, Appli
c	44	100.5	5.4	11601	2	US-08-222-617A-24	Sequence 24, Appl
c	45	99.5	5.3	3414	4	US-09-252-991A-8258	Sequence 8258, Ap

ALIGNMENTS

```
RESULT 1
US-08-827-190-8
; Sequence 8, Application US/08827190
; Patent No. 5858367
;  GENERAL INFORMATION:
;    APPLICANT:  Rather, Philip N.
;    TITLE OF INVENTION:  Methods For Screening For Antimicrobials
;    TITLE OF INVENTION:  Utilizing aarC And Compositions Thereof
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Medlen & Carroll, LLP
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/827,190
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Carroll, Peter G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  CASE-02443
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1010 base pairs
;      TYPE:  nucleic acid
```

```
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-08-827-190-8

Alignment Scores:
Pred. No.:              1.43e-185       Length:         1010
Score:                  1691.00         Matches:        336
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            90.62%          Indels:         0
DB:                     2               Gaps:           0

US-09-921-992-78 (1-372) x US-08-827-190-8 (1-1010)

Qy          1 MetHisAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCATAACCAGGCTCCAATTCAACGTAGAAAATCAACACGTATTTACGTTGGGAATGTG 60

Qy         21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrThrAsp 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCGATTGGCGATGGTGCTCCCATCGCCGTACAGTCCATGACCAATACGCGTACGACAGAC 120

Qy         41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GTCGAAGCAACGGTCAATCAAATCAAGGCGCTGGAACGCGTTGGCGCTGATATCGTCCGT 180

Qy         61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsn 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GTATCCGTACCGACGATGGACGCGGCAGAAGCGTTCAAACTCATCAAACAGCAGGTTAAC 240

Qy         81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTGCCGCTGGTGGCTGACATCCACTTCGACTATCGCATTGCGCTGAAAGTAGCGGAATAC 300

Qy        101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMet 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGCGTCGATTGTCTGCGTATTAACCCTGGCAATATCGGTAATGAAGAGCGTATTCGCATG 360

Qy        121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GTGGTTGACTGTGCGCGCGATAAAAACATTCCGATCCGTATTGGCGTTAACGCCGGATCG 420

Qy        141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSer 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTGGAAAAAGATCTGCAAGAAAAGTATGGCGAACCGACGCCGCAGGCGTTGCTGGAATCT 480

Qy        161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GCCATGCGTCATGTTGATCATCTCGATCGCCTGAACTTCGATCAGTTCAAAGTCAGCGTG 540

Qy        181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AAAGCGTCTGACGTCTTCCTCGCTGTTGAGTCTTATCGTTTGCTGGCAAAACAGATCGAT 600

Qy        201 GlnProLeuHisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSer 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CAGCCGTTGCATCTGGGGATCACCGAAGCCGGTGGTGCGCGCAGCGGGGCAGTAAAATCC 660

Qy        221 AlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GCCATTGGTTTAGGTCTGCTGCTGTCTGAAGGCATCGGCGACACGCTGCGCGTATCGCTG 720

Qy        241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GCGGCCGATCCGGTCGAAGAGATCAAAGTCGGTTTCGATATTTTGAAATCGCTGCGTATC 780

Qy        261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspVal 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CGTTCGCGAGGGATCAACTTCATCGCCTGCCCGACCTGTTCGCGTCAGGAATTTGATGTT 840

Qy        281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 ATCGGTACGGTTAACGCGCTGGAGCAACGCCTGGAAGATATCATCACTCCGATGGACGTT 900

Qy        301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TCGATTATCGGCTGCGTGGTGAATGGCCCAGGTGAGGCGCTGGTTTCTACACTCGGCGTC 960

Qy        321 ThrGlyGlyAsnLysLysSerGlyLeuTyrGluAspGlyValArgLys 336
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ACCGGCGGCAACAAGAAAAGCGGCCTCTATGAAGATGGCGTGCGCAAA 1008

RESULT 2
US-09-170-187-8
; Sequence 8, Application US/09170187
; Patent No. 6383745
;  GENERAL INFORMATION:
;    APPLICANT:  Rather, Philip N.
;    TITLE OF INVENTION:  Methods For Screening For Antimicrobials
;    TITLE OF INVENTION:  Utilizing aarC And Compositions Thereof
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Medlen & Carroll, LLP
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/170,187
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/827,190
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Carroll, Peter G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  CASE-02443
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1010 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-09-170-187-8

Alignment Scores:
Pred. No.:              1.43e-185       Length:         1010
Score:                  1691.00         Matches:        336
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            90.62%          Indels:         0
DB:                     4               Gaps:           0

US-09-921-992-78 (1-372) x US-09-170-187-8 (1-1010)

Qy          1 MetHisAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCATAACCAGGCTCCAATTCAACGTAGAAAATCAACACGTATTTACGTTGGGAATGTG 60

Qy         21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrThrAsp 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCGATTGGCGATGGTGCTCCCATCGCCGTACAGTCCATGACCAATACGCGTACGACAGAC 120

Qy         41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60
```

```
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 GTCGAAGCAACGGTCAATCAAATCAAGGCGCTGGAACGCGTTGGCGCTGATATCGTCCGT 180

Qy          61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsn 80
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 GTATCCGTACCGACGATGGACGCGGCAGAAGCGTTCAAACTCATCAAACAGCAGGTTAAC 240

Qy          81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 GTGCCGCTGGTGGCTGACATCCACTTCGACTATCGCATTGCGCTGAAAGTAGCGGAATAC 300

Qy         101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMet 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 GGCGTCGATTGTCTGCGTATTAACCCTGGCAATATCGGTAATGAAGAGCGTATTCGCATG 360

Qy         121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 GTGGTTGACTGTGCGCGCGATAAAAACATTCCGATCCGTATTGGCGTTAACGCCGGATCG 420

Qy         141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSer 160
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         421 CTGGAAAAAGATCTGCAAGAAAAGTATGGCGAACCGACGCCGCAGGCGTTGCTGGAATCT 480

Qy         161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         481 GCCATGCGTCATGTTGATCATCTCGATCGCCTGAACTTCGATCAGTTCAAAGTCAGCGTG 540

Qy         181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         541 AAAGCGTCTGACGTCTTCCTCGCTGTTGAGTCTTATCGTTTGCTGGCAAAACAGATCGAT 600

Qy         201 GlnProLeuHisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSer 220
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         601 CAGCCGTTGCATCTGGGGATCACCGAAGCCGGTGGTGCGCGCAGCGGGGCAGTAAAATCC 660

Qy         221 AlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         661 GCCATTGGTTTAGGTCTGCTGCTGTCTGAAGGCATCGGCGACACGCTGCGCGTATCGCTG 720

Qy         241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         721 GCGGCCGATCCGGTCGAAGAGATCAAAGTCGGTTTCGATATTTTGAAATCGCTGCGTATC 780

Qy         261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspVal 280
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         781 CGTTCGCGAGGGATCAACTTCATCGCCTGCCCGACCTGTTCGCGTCAGGAATTTGATGTT 840

Qy         281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         841 ATCGGTACGGTTAACGCGCTGGAGCAACGCCTGGAAGATATCATCACTCCGATGGACGTT 900

Qy         301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         901 TCGATTATCGGCTGCGTGGTGAATGGCCCAGGTGAGGCGCTGGTTTCTACACTCGGCGTC 960

Qy         321 ThrGlyGlyAsnLysLysSerGlyLeuTyrGluAspGlyValArgLys 336
               ||||||||||||||||||||||||||||||||||||||||||||||||
Db         961 ACCGGCGGCAACAAGAAAAGCGGCCTCTATGAAGATGGCGTGCGCAAA 1008
```

RESULT 3
US-08-827-190-4

```
; Sequence 4, Application US/08827190
; Patent No. 5858367
;  GENERAL INFORMATION:
;    APPLICANT:  Rather, Philip N.
;    TITLE OF INVENTION:  Methods For Screening For Antimicrobials
;    TITLE OF INVENTION:  Utilizing aarC And Compositions Thereof
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Medlen & Carroll, LLP
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/827,190
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Carroll, Peter G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  CASE-02443
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1345 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  175..1272
;      OTHER INFORMATION:  /gene= "aarC"
;      OTHER INFORMATION:  /note= "Similar to E. coli GcpE protein listed by GenBank.
US-08-827-190-4

Alignment Scores:
Pred. No.:              1.02e-177    Length:        1345
Score:                  1624.50      Matches:       325
Percent Similarity:     94.09%       Conservative:  25
Best Local Similarity:  87.37%       Mismatches:    21
Query Match:            87.06%       Indels:        2
DB:                     2            Gaps:          1

US-09-921-992-78 (1-372) x US-08-827-190-4 (1-1345)

Qy           1 MetHisAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnVal 20
               ||||||||:::::|||||||:::|||||||||||||||||||||||||||||||||||||
Db         175 ATGCATAATGAATCACCGATAAAAAGACGTAAATCCACCCGAATTTATGTAGGTAACGTG 234

Qy          21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrThrAsp 40
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         235 CCTATTGGCGATGGTGCTCCCATTGCTGTCCAATCTATGACGAATACGCGCACGACGGAT 294

Qy          41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60
               ||||||||||||||   |||||:::::||||||||||||||   ||||||||||||
Db         295 GTTGAAGCCACTGTGCGGCAAATCCAATCACTTGAGCGTGTAGGTGTTGATATCGTCCGC 354

Qy          61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsn 80
               ||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db         355 GTGTCTGTTCCTACGATGGATGCAGCAGAAGCCTTTAAATTAATTAAGCAGCGCGTGAAT 414

Qy          81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100
               ||||||||||||||||||||||||||||||||||||||||:::|||||||||||||||||
Db         415 GTGCCATTGGTTGCGGATATTCACTTTGACTACCGTATCGCGATGAAAGTGGCTGAATAT 474

Qy         101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMet 120
               ||||||||||||||||||||||||||||||||||||||:::|||||||||||||||||
Db         475 GGTGTTGACTGCCTACGAATTAACCCAGGTAATATCGGCAGTGAAGAGCGTATTCGCCAA 534

Qy         121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140
               |||||||||   ||||||    |||||||||||||||||||||||||||   ||||||
Db         535 GTCGTTGATAGTGCTCGTCATCACAACATTCCTATCCGTATAGGGGTCAATGGCGGGTCA 594

Qy         141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSer 160
               |||||||||||:::||||||||||||||||||||||||:::|||||||:::||||||
Db         595 CTGGAAAAAGATATCCAAGAAAAATACGGTGAGCCAACACCTGAAGCATTGGTTGAATCA 654
```

```
Qy        161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180
              ||||||||||||||||||   ||||||||||||||||||||||||||||||||||||||||
Db        655 GCAATGCGACATGTTGATATCTTGGACAGGCTGAATTTCGATCAGTTCAAGGTCAGTGTT 714

Qy        181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200
              |||||||||||||||||||||||||   ||||||||||||||||||:::::||||||
Db        715 AAAGCGTCGGATGTCTTTCTTGCCGTCGGCTCTTATCGTTTATTGGCGCAAAAAATTGAT 774

Qy        201 GlnProLeuHisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSer 220
              ||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db        775 CAACCACTTCACCTCGGTATTACAGAAGCGGGTGGGGCTCGTTCTGGTTCAGTGAAATCA 834

Qy        221 AlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240
              ||||||||||||||:::||||||:::|||||||||||||||||||||||:::||||||
Db        835 GCAATTGGTCTTGGTATGTTGTTGGCTGAAGGTATCGGCGATACGTTACGTATCTCACTC 894

Qy        241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260
              ||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||
Db        895 GCGGCAGATCCTGTTGAGGAAGTGAAAGTCGGTTTTGATATTCTAAAATCGTTACGGATC 954

Qy        261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspVal 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        955 CGCTCACGTGGCATCAACTTTATTGCTTGCCCAACCTGTTCACGCCAAGAATTTGATGTG 1014

Qy        281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1015 ATTGGTACGGTAAATGCTTTGGAGCAGCGCCTCGAAGATATTATCACGCCGATGGATGTC 1074

Qy        301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320
              ||||||||||||||||||||||||||||||||||   ||||||||||||||||||||
Db       1075 TCTATTATTGGTTGTGTAGTGAATGGCCCGGGTGAAGCCGAGGTTTCTACTTTAGGTGTG 1134

Qy        321 ThrGlyGlyAsnLysLysSerGlyLeuTyrGluAspGlyValArg---LysAspArgLeu 339
                  |||         |||||||||   ||||||||||||||||||   |||:::|||
Db       1135 GCTGGCGCGAAAACCAAAAGTGGTTTCTATGAAGATGGCGTTCGCAAAAAAGAGCGTTTT 1194

Qy        340 AspAsnAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeu 359
              ||||||::::::::::|||||||||||||||||||:::|||||||||||||||:::   |||
Db       1195 GATAATGACAATATTATTGATCAGCTTGAGGCGAAAATTCGCGCAAAAGCAGCAATGCTT 1254

Qy        360 AspGluAlaArgArgIleAspValGlnGlnValGlu 371
              ||||||      ||||||   :::   |||||||||
Db       1255 GATGAAAT-AACCGTATAAAGATAAACCAAGTCGAA 1289
```

```
RESULT 4
US-09-170-187-4
; Sequence 4, Application US/09170187
; Patent No. 6383745
;  GENERAL INFORMATION:
;    APPLICANT:  Rather, Philip N.
;    TITLE OF INVENTION:  Methods For Screening For Antimicrobials
;    TITLE OF INVENTION:  Utilizing aarC And Compositions Thereof
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Medlen & Carroll, LLP
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/170,187
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/827,190
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Carroll, Peter G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  CASE-02443
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1345 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  175..1272
;      OTHER INFORMATION:  /gene= "aarC"
;      OTHER INFORMATION:  /note= "Similar to E. coli GcpE protein listed by GenBank.
US-09-170-187-4

Alignment Scores:
Pred. No.:                1.02e-177       Length:          1345
Score:                    1624.50         Matches:         325
Percent Similarity:       94.09%          Conservative:    25
Best Local Similarity:    87.37%          Mismatches:      21
Query Match:              87.06%          Indels:          2
DB:                       4               Gaps:            1

US-09-921-992-78 (1-372) x US-09-170-187-4 (1-1345)

Qy          1 MetHisAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnVal 20
              |||||||||:::::|||||||:::|||||||||||||||||||||||||||||||||||
Db        175 ATGCATAATGAATCACCGATAAAAAGACGTAAATCCACCCGAATTTATGTAGGTAACGTG 234

Qy         21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrThrAsp 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        235 CCTATTGGCGATGGTGCTCCCATTGCTGTCCAATCTATGACGAATACGCGCACGACGGAT 294

Qy         41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60
              ||||||||||||||   ||||||:::::|||||||||||||||   ||||||||||||
Db        295 GTTGAAGCCACTGTGCGGCAAATCCAATCACTTGAGCGTGTAGGTGTTGATATCGTCCGC 354

Qy         61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsn 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        355 GTGTCTGTTCCTACGATGGATGCAGCAGAAGCCTTTAAATTAATTAAGCAGCGCGTGAAT 414

Qy         81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100
              ||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||||||
Db        415 GTGCCATTGGTTGCGGATATTCACTTTGACTACCGTATCGCGATGAAAGTGGCTGAATAT 474

Qy        101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMet 120
              ||||||||||||||||||||||||||||||||||||||||||:::|||||||||||||||
Db        475 GGTGTTGACTGCCTACGAATTAACCCAGGTAATATCGGCAGTGAAGAGCGTATTCGCCAA 534

Qy        121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140
              |||||||||   ||||||   ||||||||||||||||||||||||||||||   ||||||
Db        535 GTCGTTGATAGTGCTCGTCATCACAACATTCCTATCCGTATAGGGGTCAATGGCGGGTCA 594

Qy        141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSer 160
              ||||||||||:::|||||||||||||||||||||||||||||:::|||||:::||||||
Db        595 CTGGAAAAAGATATCCAAGAAAAATACGGTGAGCCAACACCTGAAGCATTGGTTGAATCA 654

Qy        161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180
              ||||||||||||||||||   ||||||||||||||||||||||||||||||||||||||||
Db        655 GCAATGCGACATGTTGATATCTTGGACAGGCTGAATTTCGATCAGTTCAAGGTCAGTGTT 714

Qy        181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200
              |||||||||||||||||||||||||   ||||||||||||||||||:::::||||||
Db        715 AAAGCGTCGGATGTCTTTCTTGCCGTCGGCTCTTATCGTTTATTGGCGCAAAAAATTGAT 774

Qy        201 GlnProLeuHisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSer 220
              ||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
```

```
Db        775 CAACCACTTCACCTCGGTATTACAGAAGCGGGTGGGGCTCGTTCTGGTTCAGTGAAATCA 834

Qy        221 AlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240

Db        835 GCAATTGGTCTTGGTATGTTGTTGGCTGAAGGTATCGGCGATACGTTACGTATCTCACTC 894

Qy        241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260

Db        895 GCGGCAGATCCTGTTGAGGAAGTGAAAGTCGGTTTTGATATTCTAAAATCGTTACGGATC 954

Qy        261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspVal 280

Db        955 CGCTCACGTGGCATCAACTTTATTGCTTGCCCAACCTGTTCACGCCAAGAATTTGATGTG 1014

Qy        281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300

Db       1015 ATTGGTACGGTAAATGCTTTGGAGCAGCGCCTCGAAGATATTATCACGCCGATGGATGTC 1074

Qy        301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320

Db       1075 TCTATTATTGGTTGTGTAGTGAATGGCCCGGGTGAAGCCGAGGTTTCTACTTTAGGTGTG 1134

Qy        321 ThrGlyGlyAsnLysLysSerGlyLeuTyrGluAspGlyValArg---LysAspArgLeu 339

Db       1135 GCTGGCGCGAAAACCAAAAGTGGTTTCTATGAAGATGGCGTTCGCAAAAAAGAGCGTTTT 1194

Qy        340 AspAsnAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeu 359

Db       1195 GATAATGACAATATTATTGATCAGCTTGAGGCGAAAATTCGCGCAAAAGCAGCAATGCTT 1254

Qy        360 AspGluAlaArgArgIleAspValGlnGlnValGlu 371

Db       1255 GATGAAAT-AACCGTATAAAGATAAACCAAGTCGAA 1289

RESULT 5
US-09-557-884-1
; Sequence 1, Application US/09557884
; Patent No. 6506581
;   GENERAL INFORMATION:
;        APPLICANT: Fleischmann et al.
;        TITLE OF INVENTION: The Nucleotide sequence of
;                            the Haemophilus influenzae Rd Genome, Fragments
;                            Thereof, and Uses Thereof
;        NUMBER OF SEQUENCES: 1
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Human Genome Sciences, Inc.
;             STREET: 9410 Key West Avenue
;             CITY: Rockville
;             STATE: MD
;             COUNTRY: USA
;             ZIP: 20850
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 3 1/2 inch diskette
;             COMPUTER: Dell Pentium
;             OPERATING SYSTEM: MS DOS v6.22
;             SOFTWARE: ASCII Text
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/557,884
;             FILING DATE: 25-Apr-2000
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/476,102
;             FILING DATE: JUN-5-1995
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Michelle S. Marks
;             REGISTRATION NUMBER: 41,971
;             REFERENCE/DOCKET NUMBER: PB186P3
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 301-309-8504
;             TELEFAX: 301-309-8439
;   INFORMATION FOR SEQ ID NO: 1:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1830121 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

Alignment Scores:
Pred. No.:              1.38e-170     Length:          1830121
Score:                  1603.50       Matches:         320
Percent Similarity:     93.66%        Conservative:    20
Best Local Similarity:  88.15%        Mismatches:      22
Query Match:            85.93%        Indels:          1
DB:                     4             Gaps:            1

US-09-921-992-78 (1-372) x US-09-557-884-1 (1-1830121)

Qy          4 GlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnValProIleGly 23

Db     390972 CAGCCAACTATTAAGCGTCGTGAATCGACAAAAATTTATGTGGGAAATGTACCAATTGGT 391031

Qy         24 AspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrThrAspValGluAla 43

Db     391032 GGGGATGCGCCTATTGCCGTGCAATCAATGACAAATACTCGCACCACTGATGTGGAAGCG 391091

Qy         44 ThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArgValSerVal 63

Db     391092 ACAGTTGCTCAAATTAAATCATTAGAACGTGTTGGTGCAGATATTGTTCGTGTATCTGTT 391151

Qy         64 ProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsnValProLeu 83

Db     391152 CCAACAATGGATGCTGCGGAAGCATTTAAACAAATTAAACAACAAGTGAATGTTCCGCTC 391211

Qy         84 ValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAsp 103

Db     391212 GTAGCAGATATTCATTTCGACTATCGTATCGCGTTAAAAGTCGCAGAATATGGAGTGGAT 391271

Qy        104 CysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMetValValAsp 123

Db     391272 TGTTTACGTATCAATCCTGGCAACATTGGTCGTGAAGATCGCGTCCGTGCCGTTGTTGAT 391331

Qy        124 CysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLys 143

Db     391332 TGTGCGCGAGACAAAAATATTCCGATTCGTATTGGTGTAAATGCAGGCTCTTTAGAAAAA 391391

Qy        144 AspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArg 163

Db     391392 GATTTGCAAGAAAAATATGGCGAACCAACGCCAGAAGCCTTGTTAGAATCCGCATTGCGT 391451

Qy        164 HisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSer 183

Db     391452 CATGTAGAAATTCTAGATCGTCTTAACTTCGATCAGTTTAAAGTGAGCGTAAAAGCCTCC 391511

Qy        184 AspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeu 203

Db     391512 GATGTATTCTTAGCGGTTGAATCTTATCGTTTACTGGCTAAAGCAATTAAACAGCCTTTA 391571

Qy        204 HisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGly 223

Db     391572 CATTTAGGCATTACAGAAGCAGGTGGCGCACGCGCTGGTGCAGTAAAATCTGCAGTGGGT 391631

Qy        224 LeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAsp 243

Db     391632 TTAGGAATGTTATTAGCTGAGGGCATTGGCGATACACTACGCGTCTCTTTGGCGGCAGAT 391691

Qy        244 ProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArg 263

Db     391692 CCTGTAGAGGAAATCAAAGTCGGTTTTGATATTTTGAAATCTTTACGGATTCGTTCAAGA 391751

Qy        264 GlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspValIleGlyThr 283

Db     391752 GGAATTAACTTTATTGCTTGCCCAACCTGTTCTCGCCAAGAATTTGATGTAATCGGTACA 391811

Qy        284 ValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspValSerIleIle 303
```

```
Db     391812 GTAAATGCGCTAGAACAACGCCTTGAAGATATTATTACACCAATGGATGTATCTATTATC 391871

Qy        304 GlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyGly 323

Db     391872 GGTTGTGTAGTGAATGGTCCTGGCGAGGCACTCGTCTCCGATCTCGGCGTAACGGGCGGT 391931

Qy        324 AsnLysLysSerGlyLeuTyrGluAspGlyValArg---LysAspArgLeuAspAsnAsn 342

Db     391932 AACAAAAAAAGCGGTTATTATCTTGACGGAGAACGCCAAAAAGAGCGTTTTGATAACGAA 391991

Qy        343 AspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAspGluAla 362

Db     391992 GATATAGTGAACCAATTAGAAGCAAAAATTCGTGCGAAAGTCGCACGACAAGATCCAAAA 392051

Qy        363 ArgArgIle 365

Db     392052 AACAGAATT 392060

RESULT 6
US-09-643-990A-1
; Sequence 1, Application US/09643990A
; Patent No. 6528289
;   GENERAL INFORMATION:
;        APPLICANT: Robert D. Fleischmann
;                   Mark D. Adams
;                   Owen White
;                   Hamilton O. Smith
;                   J. Craig Venter
;        TITLE OF INVENTION: The Nucleotide sequence of
;                            the Haemophilus influenzae Rd Genome, Fragments
;                            Thereof, and Uses Thereof
;        NUMBER OF SEQUENCES: 1
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Human Genome Sciences, Inc.
;             STREET: 9410 Key West Avenue
;             CITY: Rockville,
;             STATE: MD
;             COUNTRY: USA
;             ZIP: 20850
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 3 1/2 inch diskette
;             COMPUTER: Dell Pentium
;             OPERATING SYSTEM: MS DOS v6.22
;             SOFTWARE: ASCII Text
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/643,990A
;             FILING DATE: 23-Aug-2000
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/487,429
;             FILING DATE: 1995-06-07
;             APPLICATION NUMBER: 08/426,787
;             FILING DATE: 1995-04-21
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Kenley K. Hoover
;             REGISTRATION NUMBER: 40,302
;             REFERENCE/DOCKET NUMBER: PB186P1C1
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 301-610-5790
;             TELEFAX: 310-309-8439
;   INFORMATION FOR SEQ ID NO: 1:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1830121 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

Alignment Scores:
Pred. No.:              1.38e-170       Length:         1830121
Score:                  1603.50         Matches:        320
Percent Similarity:     93.66%          Conservative:   20
Best Local Similarity:  88.15%          Mismatches:     22
Query Match:            85.93%          Indels:         1
DB:                     4               Gaps:           1

US-09-921-992-78 (1-372) x US-09-643-990A-1 (1-1830121)

Qy          4 GlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnValProIleGly 23

Db     390972 CAGCCAACTATTAAGCGTCGTGAATCGACAAAAATTTATGTGGGAAATGTACCAATTGGT 391031

Qy         24 AspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrThrAspValGluAla 43

Db     391032 GGGGATGCGCCTATTGCCGTGCAATCAATGACAAATACTCGCACCACTGATGTGGAAGCG 391091

Qy         44 ThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArgValSerVal 63

Db     391092 ACAGTTGCTCAAATTAAATCATTAGAACGTGTTGGTGCAGATATTGTTCGTGTATCTGTT 391151

Qy         64 ProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsnValProLeu 83

Db     391152 CCAACAATGGATGCTGCGGAAGCATTTAAACAAATTAAACAACAAGTGAATGTTCCGCTC 391211

Qy         84 ValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAsp 103

Db     391212 GTAGCAGATATTCATTTCGACTATCGTATCGCGTTAAAAGTCGCAGAATATGGAGTGGAT 391271

Qy        104 CysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMetValValAsp 123

Db     391272 TGTTTACGTATCAATCCTGGCAACATTGGTCGTGAAGATCGCGTCCGTGCCGTTGTTGAT 391331

Qy        124 CysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLys 143

Db     391332 TGTGCGCGAGACAAAAATATTCCGATTCGTATTGGTGTAAATGCAGGCTCTTTAGAAAAA 391391

Qy        144 AspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArg 163

Db     391392 GATTTGCAAGAAAAATATGGCGAACCAACGCCAGAAGCCTTGTTAGAATCCGCATTGCGT 391451

Qy        164 HisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSer 183

Db     391452 CATGTAGAAATTCTAGATCGTCTTAACTTCGATCAGTTTAAAGTGAGCGTAAAAGCCTCC 391511

Qy        184 AspValPheLeuAlaValGluSerTyrArgLeuAlaLysGlnIleAspGlnProLeu 203

Db     391512 GATGTATTCTTAGCGGTTGAATCTTATCGTTTACTGGCTAAAGCAATTAAACAGCCTTTA 391571

Qy        204 HisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGly 223

Db     391572 CATTTAGGCATTACAGAAGCAGGTGGCGCACGCGCTGGTGCAGTAAAATCTGCAGTGGGT 391631

Qy        224 LeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAsp 243

Db     391632 TTAGGAATGTTATTAGCTGAGGGCATTGGCGATACACTACGCGTCTCTTTGGCGGCAGAT 391691

Qy        244 ProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArg 263

Db     391692 CCTGTAGAGGAAATCAAAGTCGGTTTTGATATTTTGAAATCTTTACGGATTCGTTCAAGA 391751

Qy        264 GlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspValIleGlyThr 283

Db     391752 GGAATTAACTTTATTGCTTGCCCAACCTGTTCTCGCCAAGAATTTGATGTAATCGGTACA 391811

Qy        284 ValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspValSerIleIle 303

Db     391812 GTAAATGCGCTAGAACAACGCCTTGAAGATATTATTACACCAATGGATGTATCTATTATC 391871

Qy        304 GlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyGly 323

Db     391872 GGTTGTGTAGTGAATGGTCCTGGCGAGGCACTCGTCTCCGATCTCGGCGTAACGGGCGGT 391931

Qy        324 AsnLysLysSerGlyLeuTyrGluAspGlyValArg---LysAspArgLeuAspAsnAsn 342
```

```
Db     391932 AACAAAAAAAGCGGTTATTATCTTGACGGAGAACGCCAAAAAGAGCGTTTTGATAACGAA 391991

Qy        343 AspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAspGluAla 362
              |||:::::::::|||||||||||||:::|||||||||||||   ::::::   |||
Db     391992 GATATAGTGAACCAATTAGAAGCAAAAATTCGTGCGAAAGTCGCACGACAAGATCCAAAA 392051

Qy        363 ArgArgIle 365
              ||||||
Db     392052 AACAGAATT 392060

RESULT 7
US-08-827-190-7
; Sequence 7, Application US/08827190
; Patent No. 5858367
;  GENERAL INFORMATION:
;    APPLICANT:  Rather, Philip N.
;    TITLE OF INVENTION:  Methods For Screening For Antimicrobials
;    TITLE OF INVENTION:  Utilizing aarC And Compositions Thereof
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Medlen & Carroll, LLP
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/827,190
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Carroll, Peter G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  CASE-02443
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1079 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-08-827-190-7

Alignment Scores:
Pred. No.:                2.76e-174      Length:          1079
Score:                    1593.50        Matches:         317
Percent Similarity:       94.93%         Conservative:    20
Best Local Similarity:    89.30%         Mismatches:      17
Query Match:              85.40%         Indels:          1
DB:                       2              Gaps:            1

US-09-921-992-78 (1-372) x US-08-827-190-7 (1-1079)

Qy          7 IleGlnArgArgLysSerThrArgIleTyrValGlyAsnValProIleGlyAspGlyAla 26
              |||:::||||||:::|||||||:::|||||||||||||||||||||||||||   |||
Db          6 ATTAAGCGTCGTGAATCGACAAAAATTTATGTGGGAAATGTACCAATTGGTGGGGATGCG 65

Qy         27 ProIleAlaValGlnSerMetThrAsnThrArgThrThrAspValGluAlaThrValAsn 46
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 CCTATTGCCGTGCAATCAATGACAAATACTCGCACCACTGATGTGGAAGCGACAGTTGCT 125

Qy         47 GlnIleLysAlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMet 66
              |||||||||:::|||||||||||||||||||||||||||||||||||||||||||||||
Db        126 CAAATTAAATCATTAGAACGTGTTGGTGCAGATATTGTTCGTGTATCTGTTCCAACAATG 185

Qy         67 AspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsnValProLeuValAlaAsp 86
              ||||||||||||||||||||||   ||||||||||||||||||||||||||||||||||
Db        186 GATGCTGCGGAAGCATTTAAACAAATTAAACAACAAGTGAATGTTCCGCTCGTAGCAGAT 245

Qy         87 IleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArg 106
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        246 ATTCATTTCGACTATCGTATCGCGTTAAAAGTCGCAGAATATGGAGTGGATTGTTTACGT 305

Qy        107 IleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMetValValAspCysAlaArg 126
              |||||||||||||||||||   |||:::|||:::|||   ||||||||||||||||||
Db        306 ATCAATCCTGGCAACATTGGTCGTGAAGATCGCGTCCGTGCCGTTGTTGATTGTGCGCGA 365

Qy        127 AspLysAsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGln 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        366 GACAAAAATATTCCGATTCGTATTGGTGTAAATGCAGGCTCTTTAGAAAAAGATTTGCAA 425

Qy        147 GluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAsp 166
              ||||||||||||||||||||||||:::|||||||||||||||||:::|||||||||:::
Db        426 GAAAAATATGGCGAACCAACGCCAGAAGCCTTGTTAGAATCCGCATTGCGTCATGTAGAA 485

Qy        167 HisLeuAspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspValPhe 186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        486 ATTCTAGATCGTCTTAACTTCGATCAGTTTAAAGTGAGCGTAAAAGCCTCCGATGTATTC 545

Qy        187 LeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeuGly 206
              ||||||||||||||||||||||||||||||   |||   ||||||||||||||||||||
Db        546 TTAGCGGTTGAATCTTATCGTTTACTGGCTAAAGCAATTAAACAGCCTTTACATTTAGGC 605

Qy        207 IleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeu 226
              ||||||||||||||||||||||||||:::|||||||||||||||||:::|||||||:::
Db        606 ATTACAGAAGCAGGTGGCGCACGCGCTGGTGCAGTAAAATCTGCAGTGGGTTTAGGAATG 665

Qy        227 LeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGlu 246
              ||||||:::||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        666 TTATTAGCTGAGGGCATTGGCGATACACTACGCGTCTCTTTGGCGGCAGATCCTGTAGAG 725

Qy        247 GluIleLysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsn 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        726 GAAATCAAAGTCGGTTTTGATATTTTGAAATCTTTACGGATTCGTTCAAGAGGAATTAAC 785

Qy        267 PheIleAlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsnAla 286
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        786 TTTATTGCTTGCCCAACCTGTTCTCGCCAAGAATTTGATGTAATCGGTACAGTAAATGCG 845

Qy        287 LeuGluGlnArgLeuGluAspIleIleThrProMetAspValSerIleIleGlyCysVal 306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        846 CTAGAACAACGCCTTGAAGATATTATTACACCAATGGATGTATCTATTATCGGTTGTGTA 905

Qy        307 ValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyGlyAsnLysLys 326
              ||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||
Db        906 GTGAATGGTCCTGGCGAGGCACTCGTCTCCGATCTCGGCGTAACGGGCGGTAACAAAAAA 965

Qy        327 SerGlyLeuTyrGluAspGlyValArg---LysAspArgLeuAspAsnAsnAspMetIle 345
              ||||||   |||   ||||||   |||   |||:::|||   ||||||  .|||::::::
Db        966 AGCGGTTATTATCTTGACGGAGAACGCCAAAAAGAGCGTTTTGATAACGAAGATATAGTG 1025

Qy        346 AspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360
              :::|||||||||||:::|||||||||||   ::::::   |||
Db       1026 AACCAATTAGAAGCAAAAATTCGTGCGAAAGTCGCACGACAAGAT 1070

RESULT 8
US-09-170-187-7
; Sequence 7, Application US/09170187
; Patent No. 6383745
;  GENERAL INFORMATION:
;    APPLICANT:  Rather, Philip N.
;    TITLE OF INVENTION:  Methods For Screening For Antimicrobials
;    TITLE OF INVENTION:  Utilizing aarC And Compositions Thereof
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Medlen & Carroll, LLP
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
```

```
;       STATE:  California
;       COUNTRY:  United States of America
;       ZIP:  94104
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/170,187
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/827,190
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Carroll, Peter G.
;       REGISTRATION NUMBER:  32,837
;       REFERENCE/DOCKET NUMBER:  CASE-02443
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (415) 705-8410
;       TELEFAX:  (415) 397-8338
;   INFORMATION FOR SEQ ID NO:  7:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1079 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  DNA (genomic)
US-09-170-187-7

Alignment Scores:
Pred. No.:              2.76e-174       Length:         1079
Score:                  1593.50         Matches:        317
Percent Similarity:     94.93%          Conservative:   20
Best Local Similarity:  89.30%          Mismatches:     17
Query Match:            85.40%          Indels:         1
DB:                     4               Gaps:           1

US-09-921-992-78 (1-372) x US-09-170-187-7 (1-1079)

Qy          7 IleGlnArgArgLysSerThrArgIleTyrValGlyAsnValProIleGlyAspGlyAla 26
              |||:::|||||||:::|||||||:::|||||||||||||||||||||||||||||||      |||
Db          6 ATTAAGCGTCGTGAATCGACAAAAATTTATGTGGGAAATGTACCAATTGGTGGGGATGCG 65

Qy         27 ProIleAlaValGlnSerMetThrAsnThrArgThrThrAspValGluAlaThrValAsn 46
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 CCTATTGCCGTGCAATCAATGACAAATACTCGCACCACTGATGTGGAAGCGACAGTTGCT 125

Qy         47 GlnIleLysAlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMet 66
              |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db        126 CAAATTAAATCATTAGAACGTGTTGGTGCAGATATTGTTCGTGTATCTGTTCCAACAATG 185

Qy         67 AspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsnValProLeuValAlaAsp 86
              ||||||||||||||||||||    ||||||||||||||||||||||||||||||||||||
Db        186 GATGCTGCGGAAGCATTTAAACAAATTAAACAACAAGTGAATGTTCCGCTCGTAGCAGAT 245

Qy         87 IleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArg 106
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        246 ATTCATTTCGACTATCGTATCGCGTTAAAAGTCGCAGAATATGGAGTGGATTGTTTACGT 305

Qy        107 IleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMetValValAspCysAlaArg 126
              ||||||||||||||||||||    |||:::|||:::|||    ||||||||||||||||||
Db        306 ATCAATCCTGGCAACATTGGTCGTGAAGATCGCGTCCGTGCCGTTGTTGATTGTGCGCGA 365

Qy        127 AspLysAsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGln 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        366 GACAAAAATATTCCGATTCGTATTGGTGTAAATGCAGGCTCTTTAGAAAAAGATTTGCAA 425

Qy        147 GluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAsp 166
              ||||||||||||||||||||||:::|||||||||||||||||:::|||||||||||:::
Db        426 GAAAAATATGGCGAACCAACGCCAGAAGCCTTGTTAGAATCCGCATTGCGTCATGTAGAA 485

Qy        167 HisLeuAspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspValPhe 186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        486 ATTCTAGATCGTCTTAACTTCGATCAGTTTAAAGTGAGCGTAAAAGCCTCCGATGTATTC 545

Qy        187 LeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeuGly 206
              ||||||||||||||||||||||||||||||||    |||   ||||||||||||||||||
Db        546 TTAGCGGTTGAATCTTATCGTTTACTGGCTAAAGCAATTAAACAGCCTTTACATTTAGGC 605

Qy        207 IleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeu 226
              ||||||||||||||||||||||:::|||||||||||||||||||:::|||||||||:::
Db        606 ATTACAGAAGCAGGTGGCGCACGCGCTGGTGCAGTAAAATCTGCAGTGGGTTTAGGAATG 665

Qy        227 LeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGlu 246
              ||||||:::||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        666 TTATTAGCTGAGGGCATTGGCGATACACTACGCGTCTCTTTGGCGGCAGATCCTGTAGAG 725

Qy        247 GluIleLysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsn 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        726 GAAATCAAAGTCGGTTTTGATATTTTGAAATCTTTACGGATTCGTTCAAGAGGAATTAAC 785

Qy        267 PheIleAlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsnAla 286
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        786 TTTATTGCTTGCCCAACCTGTTCTCGCCAAGAATTTGATGTAATCGGTACAGTAAATGCG 845

Qy        287 LeuGluGlnArgLeuGluAspIleIleThrProMetAspValSerIleIleGlyCysVal 306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        846 CTAGAACAACGCCTTGAAGATATTATTACACCAATGGATGTATCTATTATCGGTTGTGTA 905

Qy        307 ValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyGlyAsnLysLys 326
              ||||||||||||||||||||||||||||||    ||||||||||||||||||||||||||
Db        906 GTGAATGGTCCTGGCGAGGCACTCGTCTCCGATCTCGGCGTAACGGGCGGTAACAAAAAA 965

Qy        327 SerGlyLeuTyrGluAspGlyValArg---LysAspArgLeuAspAsnAsnAspMetIle 345
              ||||||   |||   ||||||   |||   |||:::|||   ||||||   |||:::::::
Db        966 AGCGGTTATTATCTTGACGGAGAACGCCAAAAAGAGCGTTTTGATAACGAAGATATAGTG 1025

Qy        346 AspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360
              :::|||||||||||||:::|||||||||||||   ::::::   |||
Db       1026 AACCAATTAGAAGCAAAAATTCGTGCGAAAGTCGCACGACAAGAT 1070

RESULT 9
US-09-252-991A-6420
; Sequence 6420, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 6420
;   LENGTH: 1644
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6420

Alignment Scores:
Pred. No.:              1.58e-148       Length:         1644
Score:                  1372.50         Matches:        267
Percent Similarity:     86.46%          Conservative:   46
Best Local Similarity:  73.76%          Mismatches:     48
Query Match:            73.55%          Indels:         1
DB:                     4               Gaps:           1

US-09-921-992-78 (1-372) x US-09-252-991A-6420 (1-1644)

Qy          1 MetHisAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnVal 20
```

```
Db        535 ATTCATAGCGCTTCTCCGATCATTCGCCGCAAGTCTCGGAAAATCTGGGTCGGCAACGTC 594

Qy         21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrThrAsp 40

Db        595 CCGGTGGGCGGCGACGCGCCGATCGCCGTGCAGAGCATGACCAACACCGAGACCTGCGAC 654

Qy         41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60

Db        655 GTCGCTGCCACCGTCGCGCAGATCCGCCGCCTGGAAGATGCCGGCGCGGACATCGTGCGG 714

Qy         61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsn 80

Db        715 GTTTCCGTCCCCGACATGGACGCCGCCGAGGCATTCGGCAAGATCAAGCAGCAGGTCAAC 774

Qy         81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100

Db        775 GTGCCGCTGGTCGCCGACATCCACTTCGACTATCGCATCGCCCTGCGCGTCGCCGAGCTG 834

Qy        101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMet 120

Db        835 GGAGTGGACTGCCTGCGCATCAATCCGGGCAACATCGGTCGCGAGGACCGGGTCAAGGCC 894

Qy        121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140

Db        895 GTGGTCGATGCCGCGCGCGAGCGCAACATCCCGATCCGTATCGGCGTCAATGCCGGTTCG 954

Qy        141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSer 160

Db        955 CTGGAAAAGGACCTGCAGAAGAAATACGGCGAACCGACCCCGGAAGCCCTGCTCGAATCG 1014

Qy        161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180

Db       1015 GCCATGCGCCACGTCGATCATCTCGACAAGCTGGACTTCCAGAACTTCAAGGTCAGCGTC 1074

Qy        181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200

Db       1075 AAGGCCTCCGACGTCTTCATGGCCGTCGCCGCCTATCGCCTGCTGGCCAGGCAGATCGAG 1134

Qy        201 GlnProLeuHisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSer 220

Db       1135 CAGCCCCTGCACCTGGGCATCACCGAGGCCGGCGGCCTGCGCTCCGGCACGGTGAAGTCG 1194

Qy        221 AlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240

Db       1195 GCGGTGGGGCTGGGCATGCTCCTGGCCGAGGGAATCGGCGACACCATCCGGATTTCCCTG 1254

Qy        241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260

Db       1255 GCTGCCGATCCGGTCGAGGAGATCAAGGTCGGTTTCGACATCCTCAAGTCCCTGCACCTG 1314

Qy        261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspVal 280

Db       1315 CGCTCCCGTGGCATCAACTTCATCGCCTGCCCGAGCTGTTCGCGGCAGAACTTCGACGTG 1374

Qy        281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300

Db       1375 GTGAAGACCATGAACGAGCTGGAAGGGCGCCTGGAGGACCTGCTGGTGCCGATGGACGTG 1434

Qy        301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320

Db       1435 GCCGTGATCGGTTGCGTCGTCAACGGTCCGGGCGAAGCCAAGGAGGCCCATGTCGGCCTC 1494

Qy        321 ThrGlyGlyAsnLysLysSerGlyLeuTyrGluAspGlyValArgLysAspArgLeuAsp 340

Db       1495 ACCGGCGGCACTCCGAACCTG---GTGTATATCGACGGCAAGCCGTCGCAGAAACTGACC 1551

Qy        341 AsnAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360

Db       1552 AACGACAACCTGGTGGACGAGCTGGAACGGCTGATCCGCCAGAAGGCGGCCGAGAAGGCC 1611

Qy        361 GluAla 362

Db       1612 GAGGCC 1617

RESULT 10
US-09-252-991A-6131/c
; Sequence 6131, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 6131
;   LENGTH: 2202
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6131

Alignment Scores:
Pred. No.:                2.46e-148     Length:          2202
Score:                    1372.50       Matches:         267
Percent Similarity:       86.46%        Conservative:    46
Best Local Similarity:    73.76%        Mismatches:      48
Query Match:              73.55%        Indels:          1
DB:                       4             Gaps:            1

US-09-921-992-78 (1-372) x US-09-252-991A-6131 (1-2202)

Qy          1 MetHisAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnVal 20

Db       2127 ATTCATAGCGCTTCTCCGATCATTCGCCGCAAGTCTCGGAAAATCTGGGTCGGCAACGTC 2068

Qy         21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrThrAsp 40

Db       2067 CCGGTGGGCGGCGACGCGCCGATCGCCGTGCAGAGCATGACCAACACCGAGACCTGCGAC 2008

Qy         41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60

Db       2007 GTCGCTGCCACCGTCGCGCAGATCCGCCGCCTGGAAGATGCCGGCGCGGACATCGTGCGG 1948

Qy         61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsn 80

Db       1947 GTTTCCGTCCCCGACATGGACGCCGCCGAGGCATTCGGCAAGATCAAGCAGCAGGTCAAC 1888

Qy         81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100

Db       1887 GTGCCGCTGGTCGCCGACATCCACTTCGACTATCGCATCGCCCTGCGCGTCGCCGAGCTG 1828

Qy        101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMet 120

Db       1827 GGAGTGGACTGCCTGCGCATCAATCCGGGCAACATCGGTCGCGAGGACCGGGTCAAGGCC 1768

Qy        121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140

Db       1767 GTGGTCGATGCCGCGCGCGAGCGCAACATCCCGATCCGTATCGGCGTCAATGCCGGTTCG 1708

Qy        141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSer 160

Db       1707 CTGGAAAAGGACCTGCAGAAGAAATACGGCGAACCGACCCCGGAAGCCCTGCTCGAATCG 1648

Qy        161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180

Db       1647 GCCATGCGCCACGTCGATCATCTCGACAAGCTGGACTTCCAGAACTTCAAGGTCAGCGTC 1588

Qy        181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200

Db       1587 AAGGCCTCCGACGTCTTCATGGCCGTCGCCGCCTATCGCCTGCTGGCCAGGCAGATCGAG 1528
```

```
Qy        201 GlnProLeuHisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSer 220
Db       1527 CAGCCCCTGCACCTGGGCATCACCGAGGCCGGCGGCCTGCGCTCCGGCACGGTGAAGTCG 1468

Qy        221 AlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240
Db       1467 GCGGTGGGGCTGGGCATGCTCCTGGCCGAGGGAATCGGCGACACCATCCGGATTTCCCTG 1408

Qy        241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260
Db       1407 GCTGCCGATCCGGTCGAGGAGATCAAGGTCGGTTTCGACATCCTCAAGTCCCTGCACCTG 1348

Qy        261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspVal 280
Db       1347 CGCTCCCGTGGCATCAACTTCATCGCCTGCCCGAGCTGTTCGCGGCAGAACTTCGACGTG 1288

Qy        281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300
Db       1287 GTGAAGACCATGAACGAGCTGGAAGGGCGCCTGGAGGACCTGCTGGTGCCGATGGACGTG 1228

Qy        301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320
Db       1227 GCCGTGATCGGTTGCGTCGTCAACGGTCCGGGCGAAGCCAAGGAGGCCCATGTCGGCCTC 1168

Qy        321 ThrGlyGlyAsnLysLysSerGlyLeuTyrGluAspGlyValArgLysAspArgLeuAsp 340
Db       1167 ACCGGCGGCACTCCGAACCTG---GTGTATATCGACGGCAAGCCGTCGCAGAAACTGACC 1111

Qy        341 AsnAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360
Db       1110 AACGACAACCTGGTGGACGAGCTGGAACGGCTGATCCGCCAGAAGGCGGCCGAGAAGGCC 1051

Qy        361 GluAla 362
Db       1050 GAGGCC 1045

RESULT 11
US-09-252-991A-6339
; Sequence 6339, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER:  US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 6339
;   LENGTH: 2088
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6339

Alignment Scores:
Pred. No.:              2.47e-147      Length:          2088
Score:                  1363.50        Matches:         266
Percent Similarity:     86.59%         Conservative:    44
Best Local Similarity:  74.30%         Mismatches:      47
Query Match:            73.07%         Indels:          1
DB:                     4              Gaps:            1

US-09-921-992-78 (1-372) x US-09-252-991A-6339 (1-2088)

Qy          5 AlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnValProIleGlyAsp 24
Db          5 TCTCCGATCATTCGCCGCAAGTCTCGGAAAATCTGGGTCGGCAACGTCCCGGTGGGCGGC 64

Qy         25 GlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrThrAspValGluAlaThr 44
Db         65 GACGCGCCGATCGCCGTGCAGAGCATGACCAACACCGAGACCTGCGACGTCGCTGCCACC 124

Qy         45 ValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArgValSerValPro 64
Db        125 GTCGCGCAGATCCGCCGCCTGGAAGATGCCGGCGCGGACATCGTGCGGGTTTCCGTCCCC 184

Qy         65 ThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsnValProLeuVal 84
Db        185 GACATGGACGCCGCCGAGGCATTCGGCAAGATCAAGCAGCAGGTCAACGTGCCGCTGGTC 244

Qy         85 AlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCys 104
Db        245 GCCGACATCCACTTCGACTATCGCATCGCCCTGCGCGTCGCCGAGCTGGGAGTGGACTGC 304

Qy        105 LeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMetValValAspCys 124
Db        305 CTGCGCATCAATCCGGGCAACATCGGTCGCGAGGACCGGGTCAAGGCCGTGGTCGATGCC 364

Qy        125 AlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAsp 144
Db        365 GCGCGCGAGCGCAACATCCCGATCCGTATCGGCGTCAATGCCGGTTCGCTGGAAAAGGAC 424

Qy        145 LeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHis 164
Db        425 CTGCAGAAGAAATACGGCGAACCGACCCCGGAAGCCCTGCTCGAATCGGCCATGCGCCAC 484

Qy        165 ValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAsp 184
Db        485 GTCGATCATCTCGACAAGCTGGACTTCCAGAACTTCAAGGTCAGCGTCAAGGCCTCCGAC 544

Qy        185 ValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeuHis 204
Db        545 GTCTTCATGGCCGTCGCCGCCTATCGCCTGCTGGCCAGGCAGATCGAGCAGCCCCTGCAC 604

Qy        205 LeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeu 224
Db        605 CTGGGCATCACCGAGGCCGGCGGCCTGCGCTCCGGCACGGTGAAGTCGGCGGTGGGGCTG 664

Qy        225 GlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspPro 244
Db        665 GGCATGCTCCTGGCCGAGGGAATCGGCGACACCATCCGGATTTCCCTGGCTGCCGATCCG 724

Qy        245 ValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGly 264
Db        725 GTCGAGGAGATCAAGGTCGGTTTCGACATCCTCAAGTCCCTGCACCTGCGCTCCCGTGGC 784

Qy        265 IleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrVal 284
Db        785 ATCAACTTCATCGCCTGCCCGAGCTGTTCGCGGCAGAACTTCGACGTGGTGAAGACCATG 844

Qy        285 AsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspValSerIleIleGly 304
Db        845 AACGAGCTGGAAGGGCGCCTGGAGGACCTGCTGGTGCCGATGGACGTGGCCGTGATCGGT 904

Qy        305 CysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyGlyAsn 324
Db        905 TGCGTCGTCAACGGTCCGGGCGAAGCCAAGGAGGCCCATGTCGGCCTCACCGGCGGCACT 964

Qy        325 LysLysSerGlyLeuTyrGluAspGlyValArgLysAspArgLeuAspAsnAsnAspMet 344
Db        965 CCGAACCTG---GTGTATATCGACGGCAAGCCGTCGCAGAAACTGACCAACGACAACCTG 1021

Qy        345 IleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAspGluAla 362
Db       1022 GTGGACGAGCTGGAACGGCTGATCCGCCAGAAGGCGGCCGAGAAGGCCGAGGCC 1075

RESULT 12
US-08-827-190-10/c
; Sequence 10, Application US/08827190
; Patent No. 5858367
;  GENERAL INFORMATION:
```

```
;    APPLICANT:  Rather, Philip N.
;    TITLE OF INVENTION:  Methods For Screening For Antimicrobials
;    TITLE OF INVENTION:  Utilizing aarC And Compositions Thereof
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Medlen & Carroll, LLP
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/827,190
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Carroll, Peter G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  CASE-02443
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  886 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-827-190-10

Alignment Scores:
Pred. No.:                1.11e-146      Length:          886
Score:                    1353.00        Matches:         268
Percent Similarity:       96.26%         Conservative:    15
Best Local Similarity:    91.16%         Mismatches:      11
Query Match:              72.51%         Indels:          0
DB:                       2              Gaps:            0

US-09-921-992-78 (1-372) x US-08-827-190-10 (1-886)

Qy         29 AlaValGlnSerMetThrAsnThrArgThrThrAspValGluAlaThrValAsnGlnIle 48
              ||||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db        884 GCTGTCCAATCTATGACGAATACGCGCACGACGGATGTTGAAGCCACTGTGCGGCAAATC 825

Qy         49 LysAlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMetAspAla 68
              ::::::|||||||||||||||   ||||||||||||||||||||||||||||||||||||
Db        824 CAATCACTTGAGCGTGTAGGTGTTGATATCGTCCGCGTGTCTGTTCCTACGATGGATGCA 765

Qy         69 AlaGluAlaPheLysLeuIleLysGlnGlnValAsnValProLeuValAlaAspIleHis 88
              ||||||||||||||||||||||||||:::||||||||||||||||||||||||||||||
Db        764 GCAGAAGCCTTTAAATTAATTAAGCAGCGCGTGAATGTGCCATTGGTTGCGGATATTCAC 705

Qy         89 PheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArgIleAsn 108
              ||||||||||||||||:::||||||||||||||||||||||||||||||||||||||||
Db        704 TTTGACTACCGTATCGCGATGAAAGTGGCTGAATATGGTGTTGACTGCCTACGAATTAAC 645

Qy        109 ProGlyAsnIleGlyAsnGluGluArgIleArgMetValValAspCysAlaArgAspLys 128
              ||||||||||||||||:::|||||||||||||||   |||||||||   ||||||
Db        644 CCAGGTAATATCGGCAGTGAAGAGCGTATTCGCCAAGTCGTTGATAGTGCTCGTCATCAC 585

Qy        129 AsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGlnGluLys 148
              ||||||||||||||||||||||||||   ||||||||||||||||:::|||||||||
Db        584 AACATTCCTATCCGTATAGGGGTCAATGGCGGGTCACTGGAAAAAGATATCCAAGAAAAA 525

Qy        149 TyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeu 168
              ||||||||||||||||:::||||||:::|||||||||||||||||||||||   |||
Db        524 TACGGTGAGCCAACACCTGAAGCATTGGTTGAATCAGCAATGCGACATGTTGATATCTTG 465

Qy        169 AspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAla 188
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        464 GACAGGCTGAATTTCGATCAGTTCAAGGTCAGTGTTAAAGCGTCGGATGTCTTTCTTGCC 405

Qy        189 ValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeuGlyIleThr 208
              |||   ||||||||||||||||||:::::|||||||||||||||||||||||||||||||
Db        404 GTCGGCTCTTATCGTTTATTGGCGCAAAAAATTGATCAACCACTTCACCTCGGTATTACA 345

Qy        209 GluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeuLeuLeu 228
              ||||||||||||||||||||||||||:::|||||||||||||||||||||:::||||||
Db        344 GAAGCGGGTGGGGCTCGTTCTGGTTCAGTGAAATCAGCAATTGGTCTTGGTATGTTGTTG 285

Qy        229 SerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGluGluIle 248
              :::|||||||||||||||||||||||:::|||||||||||||||||||||||||||:::
Db        284 GCTGAAGGTATCGGCGATACGTTACGTATCTCACTCGCGGCAGATCCTGTTGAGGAAGTG 225

Qy        249 LysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsnPheIle 268
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        224 AAAGTCGGTTTTGATATTCTAAAATCGTTACGGATCCGCTCACGTGGCATCAACTTTATT 165

Qy        269 AlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsnAlaLeuGlu 288
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        164 GCTTGCCCAACCTGTTCACGCCAAGAATTTGATGTGATTGGTACGGTAAATGCTTTGGAG 105

Qy        289 GlnArgLeuGluAspIleIleThrProMetAspValSerIleIleGlyCysValValAsn 308
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        104 CAGCGCCTCGAAGATATTATCACGCCGATGGATGTCTCTATTATTGGTTGTGTAGTGAAT 45

Qy        309 GlyProGlyGluAlaLeuValSerThrLeuGlyValThrGly 322
              ||||||||||||||||   ||||||||||||||||||   |||
Db         44 GGCCCGGGTGAAGCCGAGGTTTCTACTTTAGGTGTGGCTGGC 3

RESULT 13
US-09-170-187-10/c
; Sequence 10, Application US/09170187
; Patent No. 6383745
;  GENERAL INFORMATION:
;    APPLICANT:  Rather, Philip N.
;    TITLE OF INVENTION:  Methods For Screening For Antimicrobials
;    TITLE OF INVENTION:  Utilizing aarC And Compositions Thereof
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Medlen & Carroll, LLP
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/170,187
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/827,190
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Carroll, Peter G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  CASE-02443
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  886 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
```

```
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-170-187-10

Alignment Scores:
Pred. No.:                1.11e-146      Length:          886
Score:                    1353.00        Matches:         268
Percent Similarity:       96.26%         Conservative:    15
Best Local Similarity:    91.16%         Mismatches:      11
Query Match:              72.51%         Indels:          0
DB:                       4              Gaps:            0

US-09-921-992-78 (1-372) x US-09-170-187-10 (1-886)

Qy         29 AlaValGlnSerMetThrAsnThrArgThrThrAspValGluAlaThrValAsnGlnIle 48
Db        884 GCTGTCCAATCTATGACGAATACGCGCACGACGGATGTTGAAGCCACTGTGCGGCAAATC 825

Qy         49 LysAlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMetAspAla 68
Db        824 CAATCACTTGAGCGTGTAGGTGTTGATATCGTCCGCGTGTCTGTTCCTACGATGGATGCA 765

Qy         69 AlaGluAlaPheLysLeuIleLysGlnGlnValAsnValProLeuValAlaAspIleHis 88
Db        764 GCAGAAGCCTTTAAATTAATTAAGCAGCGCGTGAATGTGCCATTGGTTGCGGATATTCAC 705

Qy         89 PheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArgIleAsn 108
Db        704 TTTGACTACCGTATCGCGATGAAAGTGGCTGAATATGGTGTTGACTGCCTACGAATTAAC 645

Qy        109 ProGlyAsnIleGlyAsnGluGluArgIleArgMetValValAspCysAlaArgAspLys 128
Db        644 CCAGGTAATATCGGCAGTGAAGAGCGTATTCGCCAAGTCGTTGATAGTGCTCGTCATCAC 585

Qy        129 AsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGlnGluLys 148
Db        584 AACATTCCTATCCGTATAGGGGTCAATGGCGGGTCACTGGAAAAAGATATCCAAGAAAAA 525

Qy        149 TyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeu 168
Db        524 TACGGTGAGCCAACACGTGAAGCATTGGTTGAATCAGCAATGCGACATGTTGATATCTTG 465

Qy        169 AspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAla 188
Db        464 GACAGGCTGAATTTCGATCAGTTCAAGGTCAGTGTTAAAGCGTCGGATGTCTTTCTTGCC 405

Qy        189 ValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeuGlyIleThr 208
Db        404 GTCGGCTCTTATCGTTTATTGGCGCAAAAAATTGATCAACCACTTCACCTCGGTATTACA 345

Qy        209 GluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeuLeuLeu 228
Db        344 GAAGCGGGTGGGGCTCGTTCTGGTTCAGTGAAATCAGCAATTGGTCTTGGTATGTTGTTG 285

Qy        229 SerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGluGluIle 248
Db        284 GCTGAAGGTATCGGCGATACGTTACGTATCTCACTCGCGGCAGATCCTGTTGAGGAAGTG 225

Qy        249 LysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsnPheIle 268
Db        224 AAAGTCGGTTTTGATATTCTAAAATCGTTACGGATCCGCTCACGTGGCATCAACTTTATT 165

Qy        269 AlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsnAlaLeuGlu 288
Db        164 GCTTGCCCAACCTGTTCACGCCAAGAATTTGATGTGATTGGTACGGTAAATGCTTTGGAG 105

Qy        289 GlnArgLeuGluAspIleIleThrProMetAspValSerIleIleGlyCysValValAsn 308
Db        104 CAGCGCCTCGAAGATATTATCACGCCGATGGATGTCTCTATTATTGGTTGTGTAGTGAAT 45

Qy        309 GlyProGlyGluAlaLeuValSerThrLeuGlyValThrGly 322
Db         44 GGCCCGGGTGAAGCCGAGGTTTCTACTTTAGGTGTGGCTGGC 3
```

```
RESULT 14
US-09-328-352-3780
; Sequence 3780, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
;  APPLICANT: Gary L. Breton et al.
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTE
;  TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC99-03PA
;  CURRENT APPLICATION NUMBER: US/09/328,352
;  CURRENT FILING DATE: 1999-06-04
;  NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 3780
;   LENGTH: 1137
;   TYPE: DNA
;   ORGANISM: Acinetobacter baumannii
US-09-328-352-3780

Alignment Scores:
Pred. No.:                2.67e-128      Length:          1137
Score:                    1195.00        Matches:         231
Percent Similarity:       81.77%         Conservative:    65
Best Local Similarity:    63.81%         Mismatches:      66
Query Match:              64.04%         Indels:          0
DB:                       4              Gaps:            0

US-09-921-992-78 (1-372) x US-09-328-352-3780 (1-1137)

Qy          6 ProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnValProIleGlyAspGly 25
Db         34 CCAATTAAACGTCGACCAACACGTAAAATCCGTGTTGGTTCGGTCTATGTCGGTGGCGAT 93

Qy         26 AlaProIleAlaValGlnSerMetThrAsnThrArgThrThrAspValGluAlaThrVal 45
Db         94 GCACCTATTAGTGTGCAAAGTATGACAAATACCGAAACTTGCGATGTTGACGCAACTGTG 153

Qy         46 AsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArgValSerValProThr 65
Db        154 GCTCAGATTGAGCGTTGTGTTGATGCAGGTGCAGATATTATGCGTGTTTCAGTGCCTTCT 213

Qy         66 MetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsnValProLeuValAla 85
Db        214 ATGGAAGCTGCTGAAGCCTTTGGTGCAATCCGTAAGCGTGTTTCAGTTCCATTAGTAGCT 273

Qy         86 AspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeu 105
Db        274 GATATCCATTTTGACCATAGAATTGCTTTGGCAGTTGCAGATTATGGTGCGGACTGCTTG 333

Qy        106 ArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMetValValAspCysAla 125
Db        334 CGTATTAACCCGGGCAATATCGGTTCAGACCAGAAAGTTCGTGAAGTTGTGGCTGCGGCA 393

Qy        126 ArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeu 145
Db        394 CGTCATCACGGTATTTCTATGCGTATTGGCGTGAATGCGGGTTCTCTAGAAAAAGATTTA 453

Qy        146 GlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisVal 165
Db        454 CAGAAAAAATATGGCGAGCCTACAGGGCAGGCACTTCTTGAGTCAGCTTTACGTCATATT 513

Qy        166 AspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspVal 185
Db        514 GATATTTTAGACCGTCTTGACTTCCATGAGTTTAAAGTCAGTGTAAAAGCATCAAATGTG 573

Qy        186 PheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeu 205
Db        574 TTTTTAACCATGGATGCTTATCGTTTACTCTCTCAACAAATTGATAATCCATTACACCTT 633

Qy        206 GlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGly 225
Db        634 GGAGTGACTGAAGCTGGTATTTACCGTACAGGTACTGTGAAATCAGCGATTGCTCTTGGT 693
```

```
Qy        226 LeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProVal 245
Db        694 GGGTTATTGATGGAAGGCATTGGCGATACGATGCGTATTTCGCTTGCTGCTGAACCTGAA 753

Qy        246 GluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIle 265
Db        754 GATGAAATCAAGATCGGTTTTGATATCTTAAAATCGCTTGGCCTACGTTCTAACGGTATC 813

Qy        266 AsnPheIleAlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsn 285
Db        814 AACTTTATTGCTTGTCCAAGTTGTTCTCGCCAAGAATTTAACGTGATTCAGGTGATGCAG 873

Qy        286 AlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspValSerIleIleGlyCys 305
Db        874 GCTTTAGAAGAGCGTTTAGAAGATATCCGTACACCAATGGACGTTTCGGTTATTGGCTGT 933

Qy        306 ValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyGlyAsnLys 325
Db        934 AAGGTAAATGGCCCAGGTGAAGCAAAAGAAGCAGATATCGGGGTTGTTGGGGCTGCGCCT 993

Qy        326 LysSerGlyLeuTyrGluAspGlyValArgLysAspArgLeuAspAsnAsnAspMetIle 345
Db        994 CGTTCATTGGTTTATCGTAATGGTGAGAAAAGCCATTTAATTGATACCAATCAATTGGTT 1053

Qy        346 AspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAspGluAlaArgArgIle 365
Db       1054 GATGAAATCGAAACTATGGTTCGTCAACGTGTTCAAGAGCTTGAAGAAGCTAAATCTAAA 1113

Qy        366 AspVal 367
Db       1114 GAAATT 1119

RESULT 15
US-08-827-190-9/c
; Sequence 9, Application US/08827190
; Patent No. 5858367
;  GENERAL INFORMATION:
;    APPLICANT:  Rather, Philip N.
;    TITLE OF INVENTION:  Methods For Screening For Antimicrobials
;    TITLE OF INVENTION:  Utilizing aarC And Compositions Thereof
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Medlen & Carroll, LLP
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/827,190
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Carroll, Peter G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  CASE-02443
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  886 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-08-827-190-9
```

```
Alignment Scores:
Pred. No.:                2.68e-84      Length:          886
Score:                    812.00        Matches:         158
Percent Similarity:       70.65%        Conservative:    49
Best Local Similarity:    53.92%        Mismatches:      86
Query Match:              43.52%        Indels:          0
DB:                       2             Gaps:            0

US-09-921-992-78 (1-372) x US-08-827-190-9 (1-886)

Qy         30 ValGlnSerMetThrAsnThrArgThrThrAspValGluAlaThrValAsnGlnIleLys 49
Db        881 ATCCAAAGCATGACAACAACAAAAACACATGATGTAGAAGCAACGGTTGCGGAAATTAAC 822

Qy         50 AlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMetAspAlaAla 69
Db        821 CGTTTGGCTGAAGCCGGATGCCAAATCGTTCGGGTAGCATGTCCGGATGAACGGGCGGCA 762

Qy         70 GluAlaPheLysLeuIleLysGlnGlnValAsnValProLeuValAlaAspIleHisPhe 89
Db        761 AACGCCATTGCGGATATTAAAAAGCGCATTTCCATTCCTCTCGTTGTTGACATACATTTC 702

Qy         90 AspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArgIleAsnPro 109
Db        701 GATTATAAACTTGCGTTGAAAGCCATTGAAGGCGGCGCAGATAAAATCCGAATCAACCCC 642

Qy        110 GlyAsnIleGlyAsnGluGluArgIleArgMetValValAspCysAlaArgAspLysAsn 129
Db        641 GGCAATATCGGCCGGCGCGAAAAAGTTGAAGCGGTTGTTAAAGCGGCCAAAGACAAAGGC 582

Qy        130 IleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGlnGluLysTyr 149
Db        581 ATTCCGATCAGAATCGGAGTAAACGCCGGTTCATTGGAAAAACGGATTTTAGAAAAATAC 522

Qy        150 GlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeuAsp 169
Db        521 GGTTATCCGACTGCCGATGGAATGGTAGAAAGCGCACTTCATCACATTAAAATTCTTGAG 462

Qy        170 ArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAlaVal 189
Db        461 GATCTTGATTTTCACGATATTATTGTCAGCATGAAGGCCTCTGACGTGAACCTTGCAATC 402

Qy        190 GluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeuGlyIleThrGlu 209
Db        401 GAGGCTTATGAAAAAGCAGCGAAAGCGTTTGACTACCCGCTTCACCTCGGGATCACCGAG 342

Qy        210 AlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeuLeuLeuSer 229
Db        341 TCAGGAACACTGTTTGCCGGCACAGTAAAGAGCGCAGCAGGACTCGGCGCCATTTTAAGC 282

Qy        230 GluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGluGluIleLys 249
Db        281 AAAGGCATCGGGAACACCATGCGCATTTCACTAAGCGCAGACCCTGTAGAAGAGGTAAAA 222

Qy        250 ValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsnPheIleAla 269
Db        221 GTAGCAAGGGAGCTTCTGAAATCTTTCGGCTTAGCCTCCAATGCTGCCACGCTCATCTCA 162

Qy        270 CysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsnAlaLeuGluGln 289
Db        161 TGCCCGACTTGCGGCCGTATTGAGATTGATCTAATCAGCATTGCCAATGAAGTGGAAGAG 102

Qy        290 ArgLeuGluAspIleIleThrProMetAspValSerIleIleGlyCysValValAsnGly 309
Db        101 TATATTTCTAAGATAAAAGCGCCGATTAAAGTTGCTGTTCTCGGCTGCGCTGTAAACGGA 42

Qy        310 ProGlyGluAlaLeuValSerThrLeuGlyValThrGly 322
Db         41 CCTGGTGAAGCGAGAGAAGCTGATATCGGAATCGCCGGC 3


Search completed: August 15, 2003, 02:02:49
Job time : 1575 secs
```

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 15, 2003, 01:35:19 ; Search time 326 Seconds
(without alignments)
2553.723 Million cell updates/sec

Title: US-09-921-992-78
Perfect score: 1866
Sequence: 1 MHNQAPIQRRKSTRIYVGNV.....RAKASQLDEARRIDVQQVEK 372

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1504479 seqs, 1118970152 residues

Total number of hits satisfying chosen parameters: 3008958

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09921992/runat_12082003_094917_17547/app_query.fasta_1.519
-DB=Published_Applications_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09921992_@CGN_1_1_233_@runat_12082003_094917_17547
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1866	100.0	1119	9	US-09-921-992-3	Sequence 3, Appli
2	1603.5	85.9	1830121	14	US-10-329-960-1	Sequence 1, Appli
3	1282	68.7	640681	10	US-09-790-988-1	Sequence 1, Appli
4	822	44.1	1083	10	US-09-974-300-1692	Sequence 1692, Ap
5	808	43.3	1155	14	US-10-156-761-2548	Sequence 2548, Ap
c 6	808	43.3	9025608	14	US-10-156-761-1	Sequence 1, Appli
7	794	42.6	1180	10	US-09-881-752A-195	Sequence 195, App
8	786.5	42.1	1155	14	US-10-156-761-1636	Sequence 1636, Ap
9	761	40.8	1134	10	US-09-738-626-2205	Sequence 2205, Ap
c 10	761	40.8	3309400	10	US-09-738-626-1	Sequence 1, Appli
11	751.5	40.3	1164	10	US-09-712-363-110	Sequence 110, App
12	462	24.8	2520	9	US-09-921-992-1	Sequence 1, Appli
13	452	24.2	706	14	US-10-066-543-1184	Sequence 1184, Ap
14	339	18.2	670	9	US-09-921-992-22	Sequence 22, Appl
15	326	17.5	596	9	US-09-921-992-23	Sequence 23, Appl
c 16	321	17.2	584	9	US-09-921-992-21	Sequence 21, Appl
17	292	15.6	33675	9	US-09-921-992-2	Sequence 2, Appli
18	277	14.8	705	9	US-09-921-992-34	Sequence 34, Appl
19	262	14.0	601	9	US-09-921-992-11	Sequence 11, Appl
20	259	13.9	528	9	US-09-921-992-15	Sequence 15, Appl
21	258	13.8	379	9	US-09-921-992-16	Sequence 16, Appl
22	243	13.0	511	9	US-09-921-992-36	Sequence 36, Appl
23	241	12.9	498	9	US-09-921-992-37	Sequence 37, Appl
24	230	12.3	403	9	US-09-921-992-24	Sequence 24, Appl
25	221	11.8	938	9	US-09-921-992-13	Sequence 13, Appl
26	218	11.7	395	9	US-09-921-992-18	Sequence 18, Appl
27	215.5	11.5	464	9	US-09-921-992-33	Sequence 33, Appl
28	214	11.5	432	9	US-09-921-992-14	Sequence 14, Appl
29	212	11.4	395	9	US-09-921-992-17	Sequence 17, Appl
30	210	11.3	564	9	US-09-921-992-35	Sequence 35, Appl
31	198.5	10.6	443	9	US-09-921-992-12	Sequence 12, Appl
32	192.5	10.3	293	9	US-09-921-992-25	Sequence 25, Appl
33	179	9.6	388	9	US-09-921-992-44	Sequence 44, Appl
34	174	9.3	440	9	US-09-921-992-38	Sequence 38, Appl
35	170.5	9.1	211	9	US-09-921-992-45	Sequence 45, Appl
36	166.5	8.9	456	9	US-09-921-992-26	Sequence 26, Appl
37	166	8.9	348	10	US-09-974-300-6102	Sequence 6102, Ap
38	157.5	8.4	869	9	US-09-921-992-42	Sequence 42, Appl
39	155.5	8.3	551	9	US-09-921-992-41	Sequence 41, Appl
40	155	8.3	430	9	US-09-921-992-29	Sequence 29, Appl
41	155	8.3	615	9	US-09-921-992-6	Sequence 6, Appli
c 42	154.5	8.3	619	9	US-09-921-992-27	Sequence 27, Appl
43	151	8.1	422	9	US-09-921-992-28	Sequence 28, Appl
44	150	8.0	589	9	US-09-921-992-7	Sequence 7, Appli
45	149.5	8.0	399	9	US-09-921-992-47	Sequence 47, Appl

ALIGNMENTS

RESULT 1
US-09-921-992-3
; Sequence 3, Application US/09921992
; Patent No. US20020069426A1
; GENERAL INFORMATION:
; APPLICANT: Boronat, Albert;
; APPLICANT: Campos, Narciso;
; APPLICANT: Rodriguez-Concepcion, Manuel;
; APPLICANT: Rohmer, Michel;
; APPLICANT: Seeman, Myriam;
; APPLICANT: Valentin, Henry E.;
; APPLICANT: Venkatesh, Tyamagondlu V.;
; APPLICANT: Venkatramesh, Mylavarapu
; TITLE OF INVENTION: Methyl-D-Erythritol Phosphate Pathway Genes
; FILE REFERENCE: 16516.107/35-21(51897)US
; CURRENT APPLICATION NUMBER: US/09/921,992
; CURRENT FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/223,483
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 85
; SEQ ID NO 3
; LENGTH: 1119
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:

```
;   NAME/KEY: CDS
;   LOCATION: (1)..(1119)
US-09-921-992-3

Alignment Scores:
Pred. No.:              3.17e-216       Length:         1119
Score:                  1866.00         Matches:        372
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     9               Gaps:           0

US-09-921-992-78 (1-372) x US-09-921-992-3 (1-1119)

Qy          1 MetHisAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCATAACCAGGCTCCAATTCAACGTAGAAAATCAACACGTATTTACGTTGGGAATGTG 60

Qy         21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrThrAsp 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCGATTGGCGATGGTGCTCCCATCGCCGTACAGTCCATGACCAATACGCGTACGACAGAC 120

Qy         41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GTCGAAGCAACGGTCAATCAAATCAAGGCGCTGGAACGCGTTGGCGCTGATATCGTCCGT 180

Qy         61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsn 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GTATCCGTACCGACGATGGACGCGGCAGAAGCGTTCAAACTCATCAAACAGCAGGTTAAC 240

Qy         81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTGCCGCTGGTGGCTGACATCCACTTCGACTATCGCATTGCGCTGAAAGTAGCGGAATAC 300

Qy        101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMet 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGCGTCGATTGTCTGCGTATTAACCCTGGCAATATCGGTAATGAAGAGCGTATTCGCATG 360

Qy        121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GTGGTTGACTGTGCGCGCGATAAAAACATTCCGATCCGTATTGGCGTTAACGCCGGATCG 420

Qy        141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSer 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTGGAAAAAGATCTGCAAGAAAAGTATGGCGAACCGACGCCGCAGGCGTTGCTGGAATCT 480

Qy        161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GCCATGCGTCATGTTGATCATCTCGATCGCCTGAACTTCGATCAGTTCAAAGTCAGCGTG 540

Qy        181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AAAGCGTCTGACGTCTTCCTCGCTGTTGAGTCTTATCGTTTGCTGGCAAAACAGATCGAT 600

Qy        201 GlnProLeuHisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSer 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CAGCCGTTGCATCTGGGGATCACCGAAGCCGGTGGTGCGCGCAGCGGGGCAGTAAAATCC 660

Qy        221 AlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GCCATTGGTTTAGGTCTGCTGCTGTCTGAAGGCATCGGCGACACGCTGCGCGTATCGCTG 720

Qy        241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GCGGCCGATCCGGTCGAAGAGATCAAAGTCGGTTTCGATATTTTGAAATCGCTGCGTATC 780

Qy        261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspVal 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CGTTCGCGAGGGATCAACTTCATCGCCTGCCCGACCTGTTCGCGTCAGGAATTTGATGTT 840

Qy        281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 ATCGGTACGGTTAACGCGCTGGAGCAACGCCTGGAAGATATCATCACTCCGATGGACGTT 900

Qy        301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TCGATTATCGGCTGCGTGGTGAATGGCCCAGGTGAGGCGCTGGTTTCTACACTCGGCGTC 960

Qy        321 ThrGlyGlyAsnLysLysSerGlyLeuTyrGluAspGlyValArgLysAspArgLeuAsp 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ACCGGCGGCAACAAGAAAAAGCGGCCTCTATGAAGATGGCGTGCGCAAAGACCGTCTGGAC 1020

Qy        341 AsnAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AACAACGATATGATCGACCAGCTGGAAGCACGCATTCGTGCGAAAGCCAGTCAGCTGGAC 1080

Qy        361 GluAlaArgArgIleAspValGlnGlnValGluLys 372
              ||||||||||||||||||||||||||||||||||||
Db       1081 GAAGCGCGTCGAATTGACGTTCAGCAGGTTGAAAAA 1116

RESULT 2
US-10-329-960-1
; Sequence 1, Application US/10329960
; Publication No. US20030099277A1
; GENERAL INFORMATION:
;  APPLICANT: Fleischmann et al.
;  TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, F
;  TITLE OF INVENTION:  Thereof, and Uses Thereof
;  FILE REFERENCE: PB186P1
;  CURRENT APPLICATION NUMBER: US/10/329,960
;  CURRENT FILING DATE:  2003-01-02
;  PRIOR APPLICATION NUMBER: US 09/643,990
;  PRIOR FILING DATE: 2000-08-23
;  PRIOR APPLICATION NUMBER: US 08/487,429
;  PRIOR FILING DATE: 1995-06-07
;  PRIOR APPLICATION NUMBER: US 08/426,787
;  PRIOR FILING DATE: 1995-04-21
;  NUMBER OF SEQ ID NOS: 1
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 1830121
;   TYPE: DNA
;   ORGANISM: Haemophilus influenzae
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4747)..(4747)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (9921)..(9921)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (10150)..(10150)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (29298)..(29298)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (36543)..(36543)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (36551)..(36551)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (36636)..(36636)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (40808)..(40810)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
```

```
;   NAME/KEY: misc_feature
;   LOCATION: (44416)..(44416)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (44905)..(44905)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (44975)..(44975)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (45593)..(45593)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (45732)..(45732)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (47036)..(47036)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (51334)..(51334)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (51602)..(51602)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (51786)..(51786)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (51805)..(51805)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (55369)..(55369)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (65309)..(65309)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (65313)..(65313)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (80024)..(80024)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (100091)..(100091)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (102696)..(102696)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (105121)..(105121)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (107248)..(107248)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (117136)..(117136)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (119750)..(119750)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (119924)..(119924)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (120038)..(120038)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (121344)..(121344)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (122167)..(122167)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (122336)..(122336)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (131340)..(131340)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (131360)..(131360)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (139910)..(139910)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (140398)..(140398)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (142750)..(142750)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (145058)..(145058)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (145171)..(145171)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (145942)..(145942)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (147197)..(147197)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (150841)..(150841)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (152500)..(152500)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (152530)..(152530)
```

```
Alignment Scores:
Pred. No.:                1.39e-179      Length:          1830121
Score:                    1603.50        Matches:         320
Percent Similarity:       93.66%         Conservative:    20
Best Local Similarity:    88.15%         Mismatches:      22
Query Match:              85.93%         Indels:          1
DB:                       14             Gaps:            1

US-09-921-992-78 (1-372) x US-10-329-960-1 (1-1830121)

Qy          4 GlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnValProIleGly 23
              |||      |||:::|||||||:::|||||||:::|||||||||||||||||||||||||||
Db     390972 CAGCCAACTATTAAGCGTCGTGAATCGACAAAAATTTATGTGGGAAATGTACCAATTGGT 391031

Qy         24 AspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrThrAspValGluAla 43
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     391032 GGGGATGCGCCTATTGCCGTGCAATCAATGACAAATACTCGCACCACTGATGTGGAAGCG 391091

Qy         44 ThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArgValSerVal 63
              ||||||   |||||||||:::|||||||||||||||||||||||||||||||||||||||
Db     391092 ACAGTTGCTCAAATTAAATCATTAGAACGTGTTGGTGCAGATATTGTTCGTGTATCTGTT 391151

Qy         64 ProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsnValProLeu 83
              ||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db     391152 CCAACAATGGATGCTGCGGAAGCATTTAAACAAATTAAACAACAAGTGAATGTTCCGCTC 391211

Qy         84 ValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAsp 103
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     391212 GTAGCAGATATTCATTTCGACTATCGTATCGCGTTAAAAGTCGCAGAATATGGAGTGGAT 391271

Qy        104 CysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMetValValAsp 123
              ||||||||||||||||||||||||||||   |||:::|||:::|||   |||||||||
Db     391272 TGTTTACGTATCAATCCTGGCAACATTGGTCGTGAAGATCGCGTCCGTGCCGTTGTTGAT 391331

Qy        124 CysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLys 143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     391332 TGTGCGCGAGACAAAAATATTCCGATTCGTATTGGTGTAAATGCAGGCTCTTTAGAAAAA 391391

Qy        144 AspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArg 163
              ||||||||||||||||||||||||||||||:::|||||||||||||||||||||:::|||
Db     391392 GATTTGCAAGAAAAATATGGCGAACCAACGCCAGAAGCCTTGTTAGAATCCGCATTGCGT 391451

Qy        164 HisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSer 183
              |||||:::   ||||||||||||||||||||||||||||||||||||||||||||||||
Db     391452 CATGTAGAAATTCTAGATCGTCTTAACTTCGATCAGTTTAAAGTGAGCGTAAAAGCCTCC 391511

Qy        184 AspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeu 203
              ||||||||||||||||||||||||||||||||||||||||   |||   |||||||||
Db     391512 GATGTATTCTTAGCGGTTGAATCTTATCGTTTACTGGCTAAAGCAATTAAACAGCCTTTA 391571

Qy        204 HisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGly 223
              ||||||||||||||||||||||||||||||:::|||||||||||||||||||:::|||
Db     391572 CATTTAGGCATTACAGAAGCAGGTGGCGCACGCGCTGGTGCAGTAAAATCTGCAGTGGGT 391631

Qy        224 LeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAsp 243
              ||||||:::||||||:::||||||||||||||||||||||||||||||||||||||||||
Db     391632 TTAGGAATGTTATTAGCTGAGGGCATTGGCGATACACTACGCGTCTCTTTGGCGGCAGAT 391691

Qy        244 ProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArg 263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     391692 CCTGTAGAGGAAATCAAAGTCGGTTTTGATATTTTGAAATCTTTACGGATTCGTTCAAGA 391751

Qy        264 GlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspValIleGlyThr 283
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     391752 GGAATTAACTTTATTGCTTGCCCAACCTGTTCTCGCCAAGAATTTGATGTAATCGGTACA 391811

Qy        284 ValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspValSerIleIle 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     391812 GTAAATGCGCTAGAACAACGCCTTGAAGATATTATTACACCAATGGATGTATCTATTATC 391871

Qy        304 GlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyGly 323
              ||||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db     391872 GGTTGTGTAGTGAATGGTCCTGGCGAGGCACTCGTCTCCGATCTCGGCGTAACGGGCGGT 391931

Qy        324 AsnLysLysSerGlyLeuTyrGluAspGlyValArg---LysAspArgLeuAspAsnAsn 342
              ||||||||||||||||   |||   ||||||   |||   |||:::|||   ||||||
Db     391932 AACAAAAAAAGCGGTTATTATCTTGACGGAGAACGCCAAAAAGAGCGTTTTGATAACGAA 391991

Qy        343 AspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAspGluAla 362
              |||:::::::::|||||||||||||:::||||||||||||   ::::::   |||
Db     391992 GATATAGTGAACCAATTAGAAGCAAAAATTCGTGCGAAAGTCGCACGACAAGATCCAAAA 392051

Qy        363 ArgArgIle 365
                 |||||||
Db     392052 AACAGAATT 392060

RESULT 3
US-09-790-988-1
; Sequence 1, Application US/09790988
; Patent No. US20020127687A1
; GENERAL INFORMATION:
;  APPLICANT: SHIGENOBU, SHUJI
;  APPLICANT:  WATANABE, HIDEMI
;  APPLICANT:  HATTORI, MASAHIRA
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
;  FILE REFERENCE: 081356/0159
;  CURRENT APPLICATION NUMBER: US/09/790,988
;  CURRENT FILING DATE:  2001-02-23
;  PRIOR APPLICATION NUMBER: JP2000-107160
;  PRIOR FILING DATE: 2000-04-07
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 640681
;   TYPE: DNA
;   ORGANISM: Buchnera sp.
US-09-790-988-1

Alignment Scores:
Pred. No.:                2.69e-141      Length:          640681
Score:                    1282.00        Matches:         242
Percent Similarity:       83.88%         Conservative:    65
Best Local Similarity:    66.12%         Mismatches:      57
Query Match:              68.70%         Indels:          2
DB:                       10             Gaps:            2

US-09-921-992-78 (1-372) x US-09-790-988-1 (1-640681)

Qy          1 MetHisAsnGlnAlaPro---IleGlnArgArgLysSerThrArgIleTyrValGlyAsn 19
              |||:::|||:::         |||   ||||||||||||   ||||||||||||||||
Db     314272 ATGAATAATAAGTGTAAAATTATCAATAGAAGAAAATCTGATCGTATTTATGTTGGAAAA 314331

Qy         20 ValProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrThr 39
              |||   ||||||:::   ||||||||||:::||||||||||||||||||||||||||||
Db     314332 GTCGCTATTGGCAATAATGCGCCAATATCAGTTCAATCTATGACAAATACTCGTACTACT 314391

Qy         40 AspValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleVal 59
              ::::::      |||:::||||||||      |||::::::||||||   |||||||||
Db     314392 AATATCTCTGAAACTATTAATCAAATCTTAGAGTTACAAAAAGTAGGAGTAGATATTGTT 314451

Qy         60 ArgValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnVal 79
              |||:::|||:::|||   :::   ||||||||||:::||||||   ||||||:::|||
Db     314452 CGTATTTCTATACCCAATTTAAAAGCTGCAGAATCATTCAAAGAAATAAAAAAACAAACA 314511

Qy         80 AsnValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGlu 99
              ||||||||||||:::||||||||||||||||||||||||:::||||||:::      :::
Db     314512 AATGTTCCATTGATTGCAGATATACATTTTGATTACAGATTAGCTTTACAAGCTATAAAA 314571

Qy        100 TyrGlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArg 119
              ||||||   ||||||||||||||||||||||||||||||:::   |||:::
Db     314572 TATGGTGCAGATTGTTTAAGAATTAATCCTGGGAATATTGGAAATAAAAGAAGAGTATCA 314631

Qy        120 MetValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGly 139
                 ::::::      |||:::|||:::||||||||||||||||||||||||||||||||
```

```
Db      314632 GAAATCATTTCTTACGCAAAAGATGAAAATATTCCAATTCGTATTGGTGTTAATGCTGGA 314691

Qy         140 SerLeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGlu 159

Db      314692 TCTTTAGAAAAAGATATATTAAAAAAATATAAAATACCTACTCCAGATGCATTAGTAGAA 314751

Qy         160 SerAlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSer 179

Db      314752 TCAGCTATGAGGCATATTGAATACTTTGATGCTTTAAATTTTAATCAATTTAAAGTTAGT 314811

Qy         180 ValLysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIle 199

Db      314812 GTTAAAGCGTCTGATGTATTTTTAGCTATTGAATCATATCGCATGTTAGGAAAAAAAATT 314871

Qy         200 AspGlnProLeuHisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLys 219

Db      314872 ACACAACCTTTGCATATTGGCATAACTGAATCCGGTGGTTTAAGGAATGGAACAGTTAAA 314931

Qy         220 SerAlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSer 239

Db      314932 TCATCTATAGGTATTGCTTTATTATTATTAGAAGGCATTGGAGATACAATACGAGTTTCA 314991

Qy         240 LeuAlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArg 259

Db      314992 TTAGCGGCACATCCAACTGAAGAAGTAAAAGTAGGTTATGACATTTTAAAAGTTTATCT 315051

Qy         260 IleArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAsp 279

Db      315052 TTAAGAGCAAGAGGTATTAATTTTATTGCTTGTCCTACTTGTTCTAGACAAGAATTTGAT 315111

Qy         280 ValIleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAsp 299

Db      315112 GTAATTAATACAGTAAATCAACTAGAAAAAAATCTAGAAGATATCTCGACTCCCATAGAT 315171

Qy         300 ValSerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGly 319

Db      315172 GTATCAATTATTGGTTGCGTTGTTAATGGAATAGGTGAATCTAAAATAGCAACTTTAGGT 315231

Qy         320 ValThrGlyGlyAsnLysLysSerGlyLeuTyrGluAspGlyValArg---LysAspArg 338

Db      315232 CTAGCAGGAAGTCATAAGAAAAGTGCATTTTATGAAGACGGAGTAAGACAAAAAGAAAAA 315291

Qy         339 LeuAspAsnAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGln 358

Db      315292 ATAAAAAACGAAGAAATTATAGAAAAAATGGAAATTAAAATTCGAAAAAAAATAGACAAA 315351

Qy         359 LeuAspGluAlaArgArg 364

Db      315352 TTAAATAATTCAAAAAAA 315369

RESULT 4
US-09-974-300-1692
; Sequence 1692, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
;  APPLICANT: Berka, Randy M.
;  APPLICANT:  Clausen, Ib Groth
;  TITLE OF INVENTION: Methods For Monitoring Multiple Gene
;  TITLE OF INVENTION:  Expression
;  FILE REFERENCE: 10085.500-US
;  CURRENT APPLICATION NUMBER: US/09/974,300
;  CURRENT FILING DATE:  2001-10-05
;  PRIOR APPLICATION NUMBER: 09/680,598
;  PRIOR FILING DATE: 2000-10-06
;  PRIOR APPLICATION NUMBER: 60/279,526
;  PRIOR FILING DATE: 2001-03-27
;  NUMBER OF SEQ ID NOS: 8481
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1692
;   LENGTH: 1083
;   TYPE: DNA
;   ORGANISM: Bacillus licheniformis
US-09-974-300-1692
```

Alignment Scores:

Pred. No.:	1.07e-89	Length:	1083
Score:	822.00	Matches:	162
Percent Similarity:	68.45%	Conservative:	55
Best Local Similarity:	51.10%	Mismatches:	100
Query Match:	44.05%	Indels:	0
DB:	10	Gaps:	0

US-09-921-992-78 (1-372) x US-09-974-300-1692 (1-1083)

```
Qy           9 ArgArgLysSerThrArgIleTyrValGlyAsnValProIleGlyAspGlyAlaProIle 28

Db          24 CGTACAAAAACGCGTCCCGTTAAAGTGGGACCTTTAACAATAGGCGGCAATAACGAAGTC 83

Qy          29 AlaValGlnSerMetThrAsnThrArgThrThrAspValGluAlaThrValAsnGlnIle 48

Db          84 GTCATTCAAAGCATGACAACAACGAAAACACATGACGTTGAAGCAACCGTCGCCGAAATC 143

Qy          49 LysAlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMetAspAla 68

Db         144 AACAGACTCGCGGAAGCAGGATGTCAAATCGTCCGCGTCGCCTGTCCTGATGAACGGGCT 203

Qy          69 AlaGluAlaPheLysLeuIleLysGlnGlnValAsnValProLeuValAlaAspIleHis 88

Db         204 GCCGACGCCATTCCAGAGATCAAAAAGCGGATATCCATCCCTCTTGTCGTGGATATTCAT 263

Qy          89 PheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArgIleAsn 108

Db         264 TTCAACTATAAATTGGCATTAAAAGCGATCGAAGGCGGAGCCGATAAAATCCGCATCAAT 323

Qy         109 ProGlyAsnIleGlyAsnGluGluArgIleArgMetValValAspCysAlaArgAspLys 128

Db         324 CCGGGTAACATCGGCCGCCGCGAAAAGGTTGAAGCGGTCGTCAACGCAGCGAAGGAAAAG 383

Qy         129 AsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGlnGluLys 148

Db         384 GGCATTCCGATCCGGATCGGCGTCAATGCAGGCTCTCTGGAAAAACGAATCCTTGAGAAG 443

Qy         149 TyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeu 168

Db         444 TACGGCTATCCGACAGCAGACGGCATGGTCGAAAGCGCCCTGCACCACATTAAAATTCTT 503

Qy         169 AspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAla 188

Db         504 GAGGATCTCGATTTCCACGATATCATCGTCAGCATGAAGGCGTCTGATGTAAACCTGGCG 563

Qy         189 ValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeuGlyIleThr 208

Db         564 ATTGAGGCATATGAAAAAGCGGCTAAAGCTTTCGATTATCCGCTTCATTTAGGCATCACC 623

Qy         209 GluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeuLeuLeu 228

Db         624 GAATCGGGAACGCTGTTTGCCGGTACAGTGAAAAGCGCCGCGGGTCTCGGCGCGATCCTT 683

Qy         229 SerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGluGluIle 248

Db         684 TCAAAAGGAATCGGCAATACATTGCGGATTTCTTTAAGCGCGGACCCGGTCGAAGAAGTA 743

Qy         249 LysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsnPheIle 268

Db         744 AAAGTCGCCAGGGAGCTCCTGAAATCTTTCGGGCTCGCTTCCAATGCGGCAACATTGATT 803

Qy         269 AlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsnAlaLeuGlu 288

Db         804 TCCTGCCCGACTTGCGGCCGGATCGAAATCGATTTGATCTGCATTGCCAATGAATTCGAA 863

Qy         289 GlnArgLeuGluAspIleIleThrProMetAspValSerIleIleGlyCysValValAsn 308

Db         864 GATTACATCGCGAAAATCAAGGCGCCGATCAAAGTTGCCGTTCTCGGCTGTGCCGTCAAC 923

Qy         309 GlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyGlyAsnLys 325
```

```
Db        924 GGTCCCGGAGAAGCGCGGGAAGCCGATATCCGGATCGCCGGCGCACGCCGG 974

RESULT 5
US-10-156-761-2548
; Sequence 2548, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
;  APPLICANT: OMURA, SATOSHI
;  APPLICANT:  IKEDA, HARUO
;  APPLICANT:  ISHIKAWA, JUN
;  APPLICANT:  HORIKAWA, HIROSHI
;  APPLICANT:  SHIBA, TADAYOSHI
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  APPLICANT:  HATTORI, MASAHIRA
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-262
;  CURRENT APPLICATION NUMBER: US/10/156,761
;  CURRENT FILING DATE:  2002-05-29
;  PRIOR APPLICATION NUMBER: JP 2001-204089
;  PRIOR FILING DATE: 2001-05-30
;  PRIOR APPLICATION NUMBER: JP 2001-272697
;  PRIOR FILING DATE: 2001-08-02
;  NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 2548
;   LENGTH: 1155
;   TYPE: DNA
;   ORGANISM: Streptomyces avermitilis
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1155)
US-10-156-761-2548

Alignment Scores:
Pred. No.:              5.88e-88        Length:         1155
Score:                  808.00          Matches:        164
Percent Similarity:     64.11%          Conservative:   70
Best Local Similarity:  44.93%          Mismatches:     119
Query Match:            43.30%          Indels:         12
DB:                     14              Gaps:           3

US-09-921-992-78 (1-372) x US-10-156-761-2548 (1-1155)

Qy          8 GlnArgArgLysSerThrArgIleTyrValGlyAsnValProIleGlyAspGlyAlaPro 27
              :::|||||||||||||  ::::|||  ||||||  |||  :::|||      ||||||
Db         49 GAGCGCCGGAAGAGCCGGCAGATCCAGGTCGGAACCGTGGCCGTGGGCGGAGACGCACCC 108

Qy         28 IleAlaValGlnSerMetThrAsnThrArgThrThrAspValGluAlaThrValAsnGln 47
              ::::::|||||||||||||||||  ||||||||||:::|||:::  ||||||:::  |||
Db        109 GTCTCGGTGCAGTCGATGACGACGACGCGTACGTCCGACATCGGCGCCACGCTCCAGCAG 168

Qy         48 IleLysAlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMetAsp 67
              |||      |||      |||      ||||||||||||:::  ||||||  |||
Db        169 ATCGCCGAGCTCACGGCGTCCGGCTGCCAGATCGTCCGTGTGGCCTGCCCCACGCAGGAC 228

Qy         68 AlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsnValProLeuValAlaAspIle 87
                 |||:::|||    :::|||  ::::::      :::|||:::::::||||||||
Db        229 GACGCGGACGCCCTCGCGGTCATCGCCCGCAAGTCGCAGATCCCCGTGATCGCGGACATC 288

Qy         88 HisPheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArgIle 107
              ||||||     :::          |||  |||       :::|||:::
Db        289 CACTTCCAGCCGAAGTACGTGTTCGCGGCGATCGAGGCGGGCTGCGCGGCGGTCCGGGTG 348

Qy        108 AsnProGlyAsnIleGlyAsn---GluGluArgIleArgMetValValAspCysAlaArg 126
              ||||||||||||||      :::::::::::::::   :::          |||:::
Db        349 AACCCGGGCAACATCAAGCAGTTCGACGACAAGGTCAAGGAGATCGCCAAGGCGGCGAAG 408

Qy        127 AspLysAsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGln 146
              :::          ||||||||||||||||||||||||||||||:::::::   |||
Db        409 GAGCACGGCACGCCGATCCGCATCGGCGTCAACGCGGGCTCGCTCGACCGGCGGCTGCTG 468

Qy        147 GluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAsp 166
              ||||||||||||:::  ||||||:::|||||||:::|||||||||:::
Db        469 GAGAAGTACGGCAAGGCCACCCCCGAGGCGCTCGTCGAGTCGGCGCTGTGGGAGGCGTCG 528

Qy        167 HisLeuAspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspValPhe 186
                    :::      :::|||         |||:::|||||||||   :::|||
Db        529 CTCTTCGAGGAGCACGACTTCCGGGACATCAAGATCTCGGTCAAGCACAACGACCCGGTG 588

Qy        187 LeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeuGly 206
              :::   |||   :::|||||||   ||||||  |||   |||   ||||||||||||||
Db        589 GTCATGGTCAACGCGTACCGCCAGCTGGCGGCGCAGTGCGACTACCCGCTGCACCTCGGC 648

Qy        207 IleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeu 226
              :::||||||||||||   |||      |||   :::||||||||||:::      |||
Db        649 GTGACCGAGGCCGGCCCCGCCTTCCAGGGCACGATCAAGTCGGCCGTCGCCTTCGGCGCG 708

Qy        227 LeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGlu 246
              ||||||||||||||||||||||||||:::|||||||||||||:::|||   ||||||||
Db        709 CTGCTCAGCGAGGGCATCGGCGACACGATCCGCGTCTCGCTGAGCGCGCCCCCCGTCGAG 768

Qy        247 GluIleLysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsn 266
              ||||||||||||||||      ||||||:::||||||   :::|||   |||||||:::
Db        769 GAGATCAAGGTCGGCATCCAGATCCTGGAGTCGCTGAACCTGCGCCAGCGCGGCCTGGAG 828

Qy        267 PheIleAlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsnAla 286
              ::::::|||||||:::|||   |||   :::   ||||||
Db        829 ATCGTCTCCTGCCCGTCCTGCGGACGCGCCCAGGTCGACGTCTACAAGCTGGCCGAGGAG 888

Qy        287 LeuGluGlnArgLeuGluAspIleIleThrProMetAspValSerIleIleGlyCysVal 306
              :::         ||||||   :::      |||:::   |||:::::::::||||||||
Db        889 GTCACGGCGGGTCTGGAGGGCATGGAGGTGCCCCTGCGCGTCGCGGTCATGGGCTGCGTC 948

Qy        307 ValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyGlyAsnLysLys 326
              ||||||||||||||||||||      :::   |||||||||        ||||||   |||
Db        949 GTCAACGGGCCGGGGGAGGCCCGCGAGGCCGACCTCGGCGTCGCCTCCGGCAACGGCAAG 1008

Qy        327 SerGlyLeuTyrGluAspGlyValArgLysAspArgLeuAspAsnAsnAspMetIleAsp 346
                  ::::::      |||          :::      :::   :::::::::
Db       1009 GGACAGATCTTCGTGAAGGGCGAAATCATCAAGACGGTGCCCGAGTCGAAGATCGTGGAG 1068

Qy        347 GlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAspGluAlaArgArgIleAsp 366
                 |||                                  ::::::|||||   :::|||
Db       1069 ACCCTG------------------------------ATCGAGGAGGCGATGAAGATC--- 1095

Qy        367 ValGlnGlnValGlu 371
                 :::|||:::|||
Db       1096 GCCGAGCAGATGGAG 1110

RESULT 6
US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
;  APPLICANT: OMURA, SATOSHI
;  APPLICANT:  IKEDA, HARUO
;  APPLICANT:  ISHIKAWA, JUN
;  APPLICANT:  HORIKAWA, HIROSHI
;  APPLICANT:  SHIBA, TADAYOSHI
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  APPLICANT:  HATTORI, MASAHIRA
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-262
;  CURRENT APPLICATION NUMBER: US/10/156,761
;  CURRENT FILING DATE:  2002-05-29
;  PRIOR APPLICATION NUMBER: JP 2001-204089
;  PRIOR FILING DATE: 2001-05-30
;  PRIOR APPLICATION NUMBER: JP 2001-272697
;  PRIOR FILING DATE: 2001-08-02
;  NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
;   LENGTH: 9025608
;   TYPE: DNA
;   ORGANISM: Streptomyces avermitilis
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4187715)
```

```
;  OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Alignment Scores:
Pred. No.:                4.09e-82        Length:          9025608
Score:                    808.00          Matches:         164
Percent Similarity:       64.11%          Conservative:    70
Best Local Similarity:    44.93%          Mismatches:      119
Query Match:              43.30%          Indels:          12
DB:                       14              Gaps:            3

US-09-921-992-78 (1-372) x US-10-156-761-1 (1-9025608)

Qy          8 GlnArgArgLysSerThrArgIleTyrValGlyAsnValProIleGlyAspGlyAlaPro 27
              :::|||||||||||||   :::|||   ||||||   |||   :::|||      ||||||
Db    3141566 GAGCGCCGGAAGAGCCGGCAGATCCAGGTCGGAACCGTGGCCGTGGGCGGAGACGCACCC 3141507

Qy         28 IleAlaValGlnSerMetThrAsnThrArgThrThrAspValGluAlaThrValAsnGln 47
              ::::::|||||||||||||||   ||||||||||:::|||:::   ||||||:::   |||
Db    3141506 GTCTCGGTGCAGTCGATGACGACGACGCGTACGTCCGACATCGGCGCCACGCTCCAGCAG 3141447

Qy         48 IleLysAlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMetAsp 67
              |||   |||      |||      ||||||||||||:::   ||||||   |||
Db    3141446 ATCGCCGAGCTCACGGCGTCCGGCTGCCAGATCGTCCGTGTGGCCTGCCCCACGCAGGAC 3141387

Qy         68 AlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsnValProLeuValAlaAspIle 87
                |||:::|||      :::|||   ::::::      :::|||::::::|||||||||
Db    3141386 GACGCGGACGCCCTCGCGGTCATCGCCCGCAAGTCGCAGATCCCCGTGATCGCGGACATC 3141327

Qy         88 HisPheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArgIle 107
              ||||||      :::                  |||   |||      :::|||:::
Db    3141326 CACTTCCAGCCGAAGTACGTGTTCGCGGCGATCGAGGCGGGCTGCGCGGCGGTCCGGGTG 3141267

Qy        108 AsnProGlyAsnIleGlyAsn---GluGluArgIleArgMetValValAspCysAlaArg 126
              ||||||||||||||||         :::::::::::::::   :::         |||:::
Db    3141266 AACCCGGGCAACATCAAGCAGTTCGACGACAAGGTCAAGGAGATCGCCAAGGCGGCGAAG 3141207

Qy        127 AspLysAsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGln 146
              :::         ||||||||||||||||||||||||||||||||||:::::   |||
Db    3141206 GAGCACGGCACGCCGATCCGCATCGGCGTCAACGCGGGCTCGCTCGACCGGCGGCTGCTG 3141147

Qy        147 GluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAsp 166
              ||||||||||||:::   ||||||:::|||||||:::||||||||||:::
Db    3141146 GAGAAGTACGGCAAGGCCACCCCCGAGGCGCTCGTCGAGTCGGCGCTGTGGGAGGCGTCG 3141087

Qy        167 HisLeuAspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspValPhe 186
                  :::      :::|||         |||:::|||||||||   :::|||
Db    3141086 CTCTTCGAGGAGCACGACTTCCGGGACATCAAGATCTCGGTCAAGCACAACGACCCGGTG 3141027

Qy        187 LeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeuGly 206
              :::   |||   :::||||||   ||||||   |||   |||   ||||||||||||||||
Db    3141026 GTCATGGTCAACGCGTACCGCCAGCTGGCGGCGCAGTGCGACTACCCGCTGCACCTCGGC 3140967

Qy        207 IleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeu 226
              :::||||||||||||   |||      |||   :::||||||||||:::      |||
Db    3140966 GTGACCGAGGCCGGCCCCGCCTTCCAGGGCACGATCAAGTCGGCCGTCGCCTTCGGCGCG 3140907

Qy        227 LeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGlu 246
              ||||||||||||||||||||||||||:::||||||||||||:::|||   |||||||||
Db    3140906 CTGCTCAGCGAGGGCATCGGCGACACGATCCGCGTCTCGCTGAGCGCGCCCCCCGTCGAG 3140847

Qy        247 GluIleLysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsn 266
              ||||||||||||||      ||||||::||||||   :::|||   ||||||:::
Db    3140846 GAGATCAAGGTCGGCATCCAGATCCTGGAGTCGCTGAACCTGCGCCAGCGCGGCCTGGAG 3140787

Qy        267 PheIleAlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsnAla 286
                 ::::::||||||:::|||   |||   :::   ||||||
Db    3140786 ATCGTCTCCTGCCCGTCCTGCGGACGCGCCCAGGTCGACGTCTACAAGCTGGCCGAGGAG 3140727

Qy        287 LeuGluGlnArgLeuGluAspIleIleThrProMetAspValSerIleIleGlyCysVal 306
              :::         ||||||   :::      |||:::   |||:::::::::::||||||||||
Db    3140726 GTCACGGCGGGTCTGGAGGGCATGGAGGTGCCCCTGCGCGTCGCGGTCATGGGCTGCGTC 3140667

Qy        307 ValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyGlyAsnLysLys 326
              ||||||||||||||||||||||      :::   |||||||||      ||||||   |||
Db    3140666 GTCAACGGGCCGGGGGAGGCCCGCGAGGCCGACCTCGGCGTCGCCTCCGGCAACGGCAAG 3140607

Qy        327 SerGlyLeuTyrGluAspGlyValArgLysAspArgLeuAspAsnAsnAspMetIleAsp 346
                  ::::::      |||                  :::      :::   :::::::::
Db    3140606 GGACAGATCTTCGTGAAGGGCGAAATCATCAAGACGGTGCCCCGAGTCGAAGATCGTGGAG 3140547

Qy        347 GlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAspGluAlaArgArgIleAsp 366
                 |||                                      ::::::||||||   ::::||
Db    3140546 ACCCTG----------------------------ATCGAGGAGGCGATGAAGATC--- 3140520

Qy        367 ValGlnGlnValGlu 371
                 :::|||:::|||
Db    3140519 GCCGAGCAGATGGAG 3140505

RESULT 7
US-09-881-752A-195
; Sequence 195, Application US/09881752A
; Patent No. US20020115078A1
; GENERAL INFORMATION:
;  APPLICANT: Kleanthous, Harold
;  APPLICANT:  Al-Garawi, Amal
;  APPLICANT:  Miller, Charles
;  APPLICANT:  Tomb, Jean-Francois
;  APPLICANT:  Oomen, Raymond P.
;  TITLE OF INVENTION: Identification of Polynucleotides
;  TITLE OF INVENTION:  Encoding No. US20020115078A1el Helicobacter Polypeptides in t
;  TITLE OF INVENTION:  Genome
;  FILE REFERENCE: 06132/041002
;  CURRENT APPLICATION NUMBER: US/09/881,752A
;  CURRENT FILING DATE:  2001-06-15
;  PRIOR APPLICATION NUMBER: US 08/833,457
;  PRIOR FILING DATE: 1997-04-01
;  NUMBER OF SEQ ID NOS: 370
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 195
;   LENGTH: 1180
;   TYPE: DNA
;   ORGANISM: Helicobacter pylori
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (51)...(1127)
US-09-881-752A-195

Alignment Scores:
Pred. No.:                3.02e-86        Length:          1180
Score:                    794.00          Matches:         162
Percent Similarity:       66.38%          Conservative:    73
Best Local Similarity:    45.76%          Mismatches:      117
Query Match:              42.55%          Indels:          2
DB:                       10              Gaps:            2

US-09-921-992-78 (1-372) x US-09-881-752A-195 (1-1180)

Qy          9 ArgArgLysSerThrArgIleTyrValGlyAsnValProIleGlyAspGlyAlaProIle 28
              |||   |||:::   :::|||:::::::|||   |||   ||||||      |||||||||
Db         63 AGAGTTAAGACCAAGCAAATTTTTATCGGTGGCGTGGCCATAGGGGGTGATGCTCCCATA 122

Qy         29 AlaValGlnSerMetThrAsnThrArgThrThrAspValGluAlaThrValAsnGlnIle 48
              :::   ||||||||||||   :::::|||   |||:::|||:::|||   |||||||||
Db        123 AGCACGCAAAGCATGACCTTTAGCAAAACCGCTGATATTGAAAGCACTAAAAATCAAATT 182

Qy         49 LysAlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMetAspAla 68
                 |||:::      |||||||||:::||||||||||:::|||
Db        183 GACAGACTCAAACTCGCCGGGGCCGATTTAGTGAGGGTGGCGGTGAGTAATGAAAAGGAC 242

Qy         69 AlaGluAlaPheLysLeuIleLysGlnGlnValAsnValProLeuValAlaAspIleHis 88
              |||   |||   |||   :::|||:::         :::||||||:::||||||||||||
Db        243 GCTCTAGCCTTAAAAGAATTGAAAAAAGTGTCCCCTTTGCCTTTAATCGCTGATATTCAT 302

Qy         89 PheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArgIleAsn 108
              |||   |||:::   ||||||   :::|||         ||||||   :::||||||||||
```

```
Db        303 TTCCATTATAAATTCGCTCTC---ATTGCCGCTCAAAGCGTGGATGCGATCAGGATTAAC 359

Qy        109 ProGlyAsnIleGlyAsnGluGluArgIleArgMetValValAspCysAlaArgAspLys 128

Db        360 CCCGGAAACATCGGCTCTAAAGAGAAGATCAAAGCGGTGGTTGATGCTTGTAAAGAAAAA 419

Qy        129 AsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGlnGluLys 148

Db        420 AACATTCCTATAAGAATTGGCGTGAATGCTGGGAGTTTAGAAAAGCAGTTTGATCAAAAA 479

Qy        149 TyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeu 168

Db        480 TACGGA---CCCACCCCAAAAGGCATGGTAGAAAGCGCTTTGTATAACGCCAAACTTTTA 536

Qy        169 AspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAla 188

Db        537 GAAGATTTGGATTTTACCAATTTTAAGATTTCTTTAAAAGCGAGCGATGTGATTCGCACC 596

Qy        189 ValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeuGlyIleThr 208

Db        597 ATAGAAGCTTACAGGATGCTTCGCCCTCTTGTGATCTATCCTTTCCATTTGGGGGTTACG 656

Qy        209 GluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeuLeuLeu 228

Db        657 GAGGCGGGGAATCTTTTTAGCTCCAGTATCAAATCCGCTATGGCTTTAGGGGGGCTTTTA 716

Qy        229 SerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGluGluIle 248

Db        717 ATGGAGGGCATTGGGGATACGATGCGCGTATCCATCACAGGGGAATTAGAAAATGAAATC 776

Qy        249 LysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsnPheIle 268

Db        777 AAAGTGGCCAGAGCAATTTTACGCCATAGCGGGCGGTTGAAAGAAGGGATTAATTGGATT 836

Qy        269 AlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsnAlaLeuGlu 288

Db        837 TCTTGCCCCACTTGCGGGCGCATTGAAGCCAATTTAGTGGATATGGCGATCAAGGTAGAA 896

Qy        289 GlnArgLeuGluAspIleIleThrProMetAspValSerIleIleGlyCysValValAsn 308

Db        897 AAACGCTTAAGCCACATCAAAACCCCTTTAGACATTAGCGTGATGGGTTGCGTGGTGAAT 956

Qy        309 GlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyGlyAsnLysLysSerGly 328

Db        957 GCTTTGGGTGAAGCCAAGCATGCAGACATGGCGATCGCTTTTGGGAATCGCAGCGGTTTG 1016

Qy        329 LeuTyrGluAspGlyValArgLysAspArgLeuAspAsnAsnAspMetIleAspGlnLeu 348

Db       1017 ATCATTAAAGAGGGTAAAGTCATTCACAAACTGGCTGAAAAGGATTTATTTGAAACTTTT 1076

Qy        349 GluAlaArgIleArgAlaLysAlaSerGlnLeuAspGluAla 362

Db       1077 GTGATAGAAGTGGAAAATTTAGCTAAAGAAAGAGAAAAAAGT 1118

RESULT 8
US-10-156-761-1636
; Sequence 1636, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
;  APPLICANT: OMURA, SATOSHI
;  APPLICANT:  IKEDA, HARUO
;  APPLICANT:  ISHIKAWA, JUN
;  APPLICANT:  HORIKAWA, HIROSHI
;  APPLICANT:  SHIBA, TADAYOSHI
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  APPLICANT:  HATTORI, MASAHIRA
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-262
;  CURRENT APPLICATION NUMBER: US/10/156,761
;  CURRENT FILING DATE:  2002-05-29
;  PRIOR APPLICATION NUMBER: JP 2001-204089
;  PRIOR FILING DATE: 2001-05-30
;  PRIOR APPLICATION NUMBER: JP 2001-272697
;  PRIOR FILING DATE: 2001-08-02
;  NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1636
;   LENGTH: 1155
;   TYPE: DNA
;   ORGANISM: Streptomyces avermitilis
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1155)
US-10-156-761-1636

Alignment Scores:
Pred. No.:                2.37e-85        Length:          1155
Score:                    786.50          Matches:         164
Percent Similarity:       63.69%          Conservative:    71
Best Local Similarity:    44.44%          Mismatches:      121
Query Match:              42.15%          Indels:          13
DB:                       14              Gaps:            4

US-09-921-992-78 (1-372) x US-10-156-761-1636 (1-1155)

Qy          6 ProIle---GlnArgArgLysSerThrArgIleTyrValGlyAsnValProIleGlyAsp 24

Db         43 CCGATCGCCGAACGCCGTGTCTCCCGGCAGATCCAGGTCGGACCGCTGGCGGTCGGCGGC 102

Qy         25 GlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrThrAspValGluAlaThr 44

Db        103 ACGGCCCCGGTGTCGGTGCAGTCGATGACGACGACGCGTACGTCGGACATCGGCGCCACG 162

Qy         45 ValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArgValSerValPro 64

Db        163 CTCCAGCAGATCGCGGAACTCACCGCGTCCGGCTGCCAGATCGTCCGCGTCGCCTGCCCC 222

Qy         65 ThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsnValProLeuVal 84

Db        223 ACGCAGGACGACGCGGACGCCCTCGCGGTCATCGCCCGCAAGTCGCAGATCCCCGTCGTC 282

Qy         85 AlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCys 104

Db        283 GCGGACATCCACTTCCAGCCGAAGTACGTGTTCGCGGCGATCGAGGCGGGCTGCGCGGCG 342

Qy        105 LeuArgIleAsnProGlyAsnIleGlyAsn---GluGluArgIleArgMetValValAsp 123

Db        343 GTCCGGGTGAACCCGGGCAACATCAAGCAGTTCGACGACAAGGTCAAGGAGATCGCGCGC 402

Qy        124 CysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLys 143

Db        403 GCCGCCAAGGAGCACGGCACGCCGATCCGCATCGGCGTCAACGCGGGCTCGCTCGACCGG 462

Qy        144 AspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArg 163

Db        463 CGGCTGCTGGAGAAGTACGGCAAGGCCACCCCCGAGGCGCTCGTCGAGTCGGCGCTGTGG 522

Qy        164 HisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSer 183

Db        523 GAGGCGTCGCTCTTCGAGGAGCACGACTTCCGGGACATCAAGATCTCGGTCAAGCACAAC 582

Qy        184 AspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeu 203

Db        583 GACCCGGTGGTCATGGTCAACGCGTACCGCCAGCTGGCGGCGCAGTGCGACTACCCGCTG 642

Qy        204 HisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGly 223

Db        643 CACCTCGGCGTGACCGAGGCCGGCCCCGCCTTCCAGGGCACGATCAAGTCGGCCGTCGCC 702

Qy        224 LeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAsp 243

Db        703 TTCGGCGCGCTGCTCAGCGAGGGCATCGGCGACACGATCCGCGTCTCGCTGAGCGCGCCC 762

Qy        244 ProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArg 263

Db        763 CCCGTCGAGGAGATCAAGGTCGGCATCCAGATCCTGGAGTCGCTGGGCCTCAGGCAGCGG 822
```

```
Qy        264 GlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspValIleGlyThr 283
Db        823 CGCCTGGAGATCGTCTCCTGCCCGTCCTGCGGGCGCGCCCAGGTCGACGTCTACAAGCTG 882

Qy        284 ValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspValSerIleIle 303
Db        883 GCCGAGGAGGTCACGGCGGGTCTGGAGGGCATGGAGGTGCCCCTGCGCGTCGCGGTCATG 942

Qy        304 GlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyGly 323
Db        943 GGCTGCGTCGTCAACGGGCCGGGGGAGGCCCGCGAGGCCGACCTCGGCGTCGCCTCCGGC 1002

Qy        324 AsnLysLysSerGlyLeuTyrGluAspGlyValArgLysAspArgLeuAspAsnAsnAsp 343
Db       1003 AACGGCAAGGGACAGATCTTCGTGAAGGGCGAGGTCATCAAGACGGTGCCCGAGTCGAAG 1062

Qy        344 MetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAspGluAlaArg 363
Db       1063 ATCGTGGAGACCCTG----------ATCGAGGAGGCGATG 1092

Qy        364 ArgIleAspValGlnGlnValGluLys 372
Db       1093 AAGATC---GCCGAGCAGATGGAGCAG 1116

RESULT 9
US-09-738-626-2205
; Sequence 2205, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
;  APPLICANT: NAKAGAWA, SATOSHI
;  APPLICANT:  MIZOGUCHI, HIROSHI
;  APPLICANT:  ANDO, SEIKO
;  APPLICANT:  HAYASHI, MIKIRO
;  APPLICANT:  OCHIAI, KEIKO
;  APPLICANT:  YOKOI, HARUHIKO
;  APPLICANT:  TATEISHI, NAOKO
;  APPLICANT:  SENOH, AKIHIRO
;  APPLICANT:  IKEDA, MASATO
;  APPLICANT:  OZAKI, AKIO
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-125
;  CURRENT APPLICATION NUMBER: US/09/738,626
;  CURRENT FILING DATE:  2000-12-18
;  PRIOR APPLICATION NUMBER: JP 99/377484
;  PRIOR FILING DATE: 1999-12-16
;  PRIOR APPLICATION NUMBER: JP 00/159162
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: JP 00/280988
;  PRIOR FILING DATE: 2000-08-03
;  NUMBER OF SEQ ID NOS: 7059
;  SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 2205
;   LENGTH: 1134
;   TYPE: DNA
;   ORGANISM: Corynebacterium glutamicum
US-09-738-626-2205

Alignment Scores:
Pred. No.:                2.85e-82        Length:          1134
Score:                    761.00          Matches:         159
Percent Similarity:       62.12%          Conservative:    64
Best Local Similarity:    44.29%          Mismatches:      124
Query Match:              40.78%          Indels:          12
DB:                       10              Gaps:            3

US-09-921-992-78 (1-372) x US-09-738-626-2205 (1-1134)

Qy          9 ArgArgLysSerThrArgIleTyrValGlyAsnValProIleGlyAspGlyAlaProIle 28
Db         13 CGTCGTAAAACACGCCAACTCATGGTCGGCAAAGTGGGCGTTGGTTCGGATCACCCGATT 72

Qy         29 AlaValGlnSerMetThrAsnThrArgThrThrAspValGluAlaThrValAsnGlnIle 48

Db         73 TCCGTCCAGTCGATGACCACCACCAAAACCCACGACATCAACGGCACCCTGCAACAGATC 132

Qy         49 LysAlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMetAspAla 68
Db        133 GCACAGTTGACAGCCACCGGTTGTGACATCGTCCGCGTTGCCTGCCCAAAGACTGTTGAT 192

Qy         69 AlaGluAlaPheLysLeuIleLysGlnGlnValAsnValProLeuValAlaAspIleHis 88
Db        193 GCGGAAGCACTGCCGATCATCGCAAAGAAGTCTCCGATCCCAGTGATCGCAGATATCCAC 252

Qy         89 PheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArgIleAsn 108
Db        253 TTCCAGCCCAAGTACATCTTCGCGGCAATCGATGCAGGTTGCGCCGCCGTTCGTGTGAAC 312

Qy        109 ProGlyAsnIleGlyAsn---GluGluArgIleArgMetValValAspCysAlaArgAsp 127
Db        313 CCAGGCAACATCAAGGAATTCGATGGTCGCGTTAAAGAAGTAGCAAAAGCTGCAGGCGAT 372

Qy        128 LysAsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGlnGlu 147
Db        373 GCCGGAATTCCAATTCGTATTGGTGTCAACGGCGGATCCCTGGATAAGCGCATCCTGGAC 432

Qy        148 LysTyr---GlyGluProThrProGlnAlaLeuGluSerAlaMetArgHisValAsp 166
Db        433 AAATACCACGGCAAAGCCACCCCAGAAGCTCTCGTGGAATCCGCAATGTGGGAAGCCGGC 492

Qy        167 HisLeuAspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspValPhe 186
Db        493 CTGTTTGAAGAGCACGGCTTCGGCGACATCGCAATCTCTGTGAAGCACTCCGACCCAGTA 552

Qy        187 LeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeuGly 206
Db        553 CTCATGGTGGAGGCCTACCGCCAGCTCGCTGAACAAAGCGACTACCCACTGCACCTCGGT 612

Qy        207 IleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeu 226
Db        613 GTTACTGAAGCTGGTCCCAAGTTCATGGGAACAATCAAGTCTTCCGTAGCATTCGGCGCT 672

Qy        227 LeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGlu 246
Db        673 CTGCTGTCCCAGGGCATCGGCGACACTATCCGTGTGTCTCTTTCTGCTGACCCAGTGGAA 732

Qy        247 GluIleLysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsn 266
Db        733 GAAATCAAGGTTGGCGACCAGATTCTGCAGTCCCTCAACCTGCGCCCACGCAAGCTGGAA 792

Qy        267 PheIleAlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsnAla 286
Db        793 ATCGTGTCCTGCCCATCATGTGGCCGCGCACAGGTCGATGTGTACTCACTTGCTGAAGAA 852

Qy        287 LeuGluGlnArgLeuGluAspIleIleThrProMetAspValSerIleIleGlyCysVal 306
Db        853 GTCACCGAAGCACTCGACGGCATGGAAGTTCCACTGCGCGTCGCTGTCATGGGTTGCGTT 912

Qy        307 ValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyGlyAsnLysLys 326
Db        913 GTTAACGGCCCAGGTGAGGCTCGCGACGCTGACCTCGGTGTTGCATCCGGTAACGGCAAG 972

Qy        327 SerGlyLeuTyrGluAspGlyValArgLysAspArgLeuAspAsnAsnAspMetIleAsp 346
Db        973 GGCCAGATCTTTGTCAAGGGCGAAGTCATCAAGACTGTCCCAGAATCCCAGATCGTGGAA 1032

Qy        347 GlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAspGluAlaArgArgIle 365
Db       1033 ACCCTC----------ATCGAAGAAGCAATGCGTATC 1059

RESULT 10
US-09-738-626-1/c
; Sequence 1, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
;  APPLICANT: NAKAGAWA, SATOSHI
;  APPLICANT:  MIZOGUCHI, HIROSHI
```

```
;  APPLICANT:  ANDO, SEIKO
;  APPLICANT:  HAYASHI, MIKIRO
;  APPLICANT:  OCHIAI, KEIKO
;  APPLICANT:  YOKOI, HARUHIKO
;  APPLICANT:  TATEISHI, NAOKO
;  APPLICANT:  SENOH, AKIHIRO
;  APPLICANT:  IKEDA, MASATO
;  APPLICANT:  OZAKI, AKIO
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-125
;  CURRENT APPLICATION NUMBER: US/09/738,626
;  CURRENT FILING DATE:  2000-12-18
;  PRIOR APPLICATION NUMBER: JP 99/377484
;  PRIOR FILING DATE: 1999-12-16
;  PRIOR APPLICATION NUMBER: JP 00/159162
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: JP 00/280988
;  PRIOR FILING DATE: 2000-08-03
;  NUMBER OF SEQ ID NOS: 7059
;  SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 1
;   LENGTH: 3309400
;   TYPE: DNA
;   ORGANISM: Corynebacterium glutamicum
US-09-738-626-1

Alignment Scores:
Pred. No.:                4.52e-77        Length:          3309400
Score:                    761.00          Matches:         159
Percent Similarity:       62.12%          Conservative:    64
Best Local Similarity:    44.29%          Mismatches:      124
Query Match:              40.78%          Indels:          12
DB:                       10              Gaps:            3

US-09-921-992-78 (1-372) x US-09-738-626-1 (1-3309400)

Qy          9 ArgArgLysSerThrArgIleTyrValGlyAsnValProIleGlyAspGlyAlaProIle 28
              ||||||||||:::   :::::   ||||||   |||   :::|||          ||||||
Db    2128471 CGTCGTAAAACACGCCAACTCATGGTCGGCAAAGTGGGCGTTGGTTCGGATCACCCGATT 2128412

Qy         29 AlaValGlnSerMetThrAsnThrArgThrThrAspValGluAlaThrValAsnGlnIle 48
              :::|||||||||||||||   |||:::|||   |||:::      |||:::   ||||||
Db    2128411 TCCGTCCAGTCGATGACCACCACCAAAACCCACGACATCAACGGCACCCTGCAACAGATC 2128352

Qy         49 LysAlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMetAspAla 68
                 |||       |||   ||||||||||||||::   |||
Db    2128351 GCACAGTTGACAGCCACCGGTTGTGACATCGTCCGCGTTGCCTGCCCAAAGACTGTTGAT 2128292

Qy         69 AlaGluAlaPheLysLeuIleLysGlnGlnValAsnValProLeuValAlaAspIleHis 88
              |||||||||     :::::   ::::|||::::::||||||||||||
Db    2128291 GCGGAAGCACTGCCGATCATCGCAAAGAAGTCTCCGATCCCAGTGATCGCAGATATCCAC 2128232

Qy         89 PheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArgIleAsn 108
              |||      :::                 :::   |||         :::|||:::|||
Db    2128231 TTCCAGCCCAAGTACATCTTCGCGGCAATCGATGCAGGTTGCGCCGCCGTTCGTGTGAAC 2128172

Qy        109 ProGlyAsnIleGlyAsn---GluGluArgIleArgMetValValAspCysAlaArgAsp 127
              ||||||||||||        :::   |||::::::   |||         |||   |||
Db    2128171 CCAGGCAACATCAAGGAATTCGATGGTCGCGTTAAAGAAGTAGCAAAAGCTGCAGGCGAT 2128112

Qy        128 LysAsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGlnGlu 147
              ||||||||||||||||||||||||   |||||||||:::|||   :::   :::
Db    2128111 GCCGGAATTCCAATTCGTATTGGTGTCAACGGCGGATCCCTGGATAAGCGCATCCTGGAC 2128052

Qy        148 LysTyr---GlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAsp 166
              ||||||   |||:::  ||||||:::||||||:::||||||||||||
Db    2128051 AAATACCACGGCAAAGCCACCCCAGAAGCTCTCGTGGAATCCGCAATGTGGGAAGCCGGC 2127992

Qy        167 HisLeuAspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspValPhe 186
                 :::       |||           :::||||||||   ||||||
Db    2127991 CTGTTTGAAGAGCACGGCTTCGGCGACATCGCAATCTCTGTGAAGCACTCCGACCCAGTA 2127932

Qy        187 LeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeuGly 206
                |||   ||||||:::||||||   ||||||:::|||   |||   ||||||||||||||||
Db    2127931 CTCATGGTGGAGGCCTACCGCCAGCTCGCTGAACAAAGCGACTACCCACTGCACCTCGGT 2127872

Qy        207 IleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeu 226
              :::||||||||||||            |||   :::||||||:::::       |||
Db    2127871 GTTACTGAAGCTGGTCCCAAGTTCATGGGAACAATCAAGTCTTCCGTAGCATTCGGCGCT 2127812

Qy        227 LeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGlu 246
              |||||||||:::||||||||||||||||:::||||||||||||:::|||||||||||||||
Db    2127811 CTGCTGTCCCAGGGCATCGGCGACACTATCCGTGTGTCTCTTTCTGCTGACCCAGTGGAA 2127752

Qy        247 GluIleLysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsn 266
              ||||||||||||||      ||||||:::||||||   :::|||   |||   :::
Db    2127751 GAAATCAAGGTTGGCGACCAGATTCTGCAGTCCCTCAACCTGCGCCCACGCAAGCTGGAA 2127692

Qy        267 PheIleAlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsnAla 286
                 ::::::|||||:::|||   |||   :::   ||||||
Db    2127691 ATCGTGTCCTGCCCATCATGTGGCCGCGCACAGGTCGATGTGTACTCACTTGCTGAAGAA 2127632

Qy        287 LeuGluGlnArgLeuGluAspIleIleThrProMetAspValSerIleIleGlyCysVal 306
              :::   :::   |||:::   :::       |||:::   |||::::::::::|||||||||
Db    2127631 GTCACCGAAGCACTCGACGGCATGGAAGTTCCACTGCGCGTCGCTGTCATGGGTTGCGTT 2127572

Qy        307 ValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyGlyAsnLysLys 326
              ||||||||||||||||||||      :::   |||||||||      ||||||   |||
Db    2127571 GTTAACGGCCCAGGTGAGGCTCGCGACGCTGACCTCGGTGTTGCATCCGGTAACGGCAAG 2127512

Qy        327 SerGlyLeuTyrGluAspGlyValArgLysAspArgLeuAspAsnAsnAspMetIleAsp 346
                    ::::::      |||                 :::       :::   :::::::::
Db    2127511 GGCCAGATCTTTGTCAAGGGCGAAGTCATCAAGACTGTCCCAGAATCCCAGATCGTGGAA 2127452

Qy        347 GlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAspGluAlaArgArgIle 365
                 |||                              ::::::||||||   ||||||
Db    2127451 ACCCTC------------------------------ATCGAAGAAGCAATGCGTATC 2127425

RESULT 11
US-09-712-363-110
; Sequence 110, Application US/09712363
; Patent No. US20020164588A1
; GENERAL INFORMATION:
;  APPLICANT: Eisenberg, David
;  APPLICANT:  Rotstein, Sergio H.
;  APPLICANT:  Marcotte, Edward M.
;  TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
;  TITLE OF INVENTION:  INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS
;  FILE REFERENCE: 07419-032001
;  CURRENT APPLICATION NUMBER: US/09/712,363
;  CURRENT FILING DATE:  2000-11-13
;  PRIOR APPLICATION NUMBER: PCT/US00/02246
;  PRIOR FILING DATE: 2000-01-28
;  PRIOR APPLICATION NUMBER: 60/179,531
;  PRIOR FILING DATE: 2000-02-01
;  PRIOR APPLICATION NUMBER: 60/117,844
;  PRIOR FILING DATE: 1999-01-29
;  PRIOR APPLICATION NUMBER: 60/118,206,
;  PRIOR FILING DATE: 1999-02-01
;  PRIOR APPLICATION NUMBER: 60/126,593
;  PRIOR FILING DATE: 1999-03-26
;  PRIOR APPLICATION NUMBER: 60/134,093
;  PRIOR FILING DATE: 1999-05-14
;  PRIOR APPLICATION NUMBER: 60/134,092
;  PRIOR FILING DATE: 1999-05-14
;  PRIOR APPLICATION NUMBER: 60/165,124
;  PRIOR FILING DATE: 1999-11-12
;  PRIOR APPLICATION NUMBER: 60/165,086
;  PRIOR FILING DATE: 1999-11-12
;  NUMBER OF SEQ ID NOS: 292
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 110
;   LENGTH: 1164
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
US-09-712-363-110
```

Alignment Scores:

Pred. No.:	4.2e-81	Length:	1164
Score:	751.50	Matches:	156
Percent Similarity:	60.27%	Conservative:	64
Best Local Similarity:	42.74%	Mismatches:	134
Query Match:	40.27%	Indels:	11
DB:	10	Gaps:	2

US-09-921-992-78 (1-372) x US-09-712-363-110 (1-1164)

```
Qy          9 ArgArgLysSerThrArgIleTyrValGlyAsnValProIleGlyAspGlyAlaProIle 28
              ||||||   :::   :::::   ||||||||||||   :::|||           ||:::
Db         52 CGGCGCGCCACCCGTCAGCTGATGGTCGGCAACGTCGGCGTGGGCAGTGACCATCCGGTC 111

Qy         29 AlaValGlnSerMetThrAsnThrArgThrThrAspValGluAlaThrValAsnGlnIle 48
              :::||||||||||||      |||:::|||   ||||||   :::|||:::   ||||||
Db        112 TCGGTGCAATCGATGTGCACCACCAAAACCCACGACGTCAACTCGACATTGCAACAAATC 171

Qy         49 LysAlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMetAspAla 68
                 |||       |||   ||||||||||||||||:::   |||      :::
Db        172 GCCGAGCTGACCGCGGCCGGATGCGACATCGTGCGGGTGGCCTGCCCGCGCCAGGAGGAC 231

Qy         69 AlaGluAlaPheLysLeuIleLysGlnGlnValAsnValProLeuValAlaAspIleHis 88
              |||:::|||      |||   :::         :::|||:::|||||||||||||||||
Db        232 GCCGACGCGCTGGCCGAGATCGCCCGGCACAGCCAGATCCCGGTAGTCGCGGACATACAT 291

Qy         89 PheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArgIleAsn 108
              |||      |||                 :::   |||       :::|||:::|||
Db        292 TTCCAGCCGCGCTACATATTCGCCGCCATCGACGCTGGATGTGCCGCGGTGCGGGTCAAC 351

Qy        109 ProGlyAsnIleGlyAsn---GluGluArgIleArgMetValValAspCysAlaArgAsp 127
              ||||||||||||      :::   |||:::      |||       |||
Db        352 CCGGGCAACATCAAGGAGTTTGACGGCCGGGTGGGTGAGGTCGCCAAGGCGGCGGGTGCG 411

Qy        128 LysAsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGlnGlu 147
              ||||||||||||||||||||||||||||||||||:::|||      |||
Db        412 GCCGGGATCCCGATCCGAATCGGTGTCAACGCCGGTTCGCTGGACAAACGGTTCATGGAG 471

Qy        148 LysTyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHis 167
              |||||||||:::   ||||||:::|||||||:::||||||||||:::
Db        472 AAGTATGGCAAAGCCACGCCCGAGGCGCTGGTTGAGTCGGCGCTGTGGGAGGCTTCGCTT 531

Qy        168 LeuAspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspValPheLeu 187
              :::      |||       |||:::||||||||   :::|||      :::
Db        532 TTCGAGGAGCATGGCTTCGGTGACATCAAGATCAGCGTCAAGCACAACGACCCGGTGGTG 591

Qy        188 AlaValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeuGlyIle 207
              |||   ::::||   |||||||||   :::   |||   ||||||||||||||:::
Db        592 ATGGTCGCCGCCTACGAGCTGCTTGCTGCACGGTGCGACTACCCACTGCACCTCGGTGTC 651

Qy        208 ThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeuLeu 227
              ||||||||||||   |||      |||   :::||||||||:::      |||   |||
Db        652 ACCGAGGCCGGCCCTGCTTTCCAGGGCACCATCAAGTCCGCGGTTGCCTTCGGCGCGTTG 711

Qy        228 LeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGluGlu 247
              ||||||.  ||||||||||||||:::||||||||||||:::|||   ||||||||||||
Db        712 CTGTCGCGGGGCATAGGCGACACCATCCGGGTGTCGTTGTCGGCCCCGCCGGTCGAGGAA 771

Qy        248 IleLysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsnPhe 267
              :::||||||||      :::|||:::||||||   :::|||   |||   :::
Db        772 GTCAAGGTGGGCAATCAGGTTCTCGAGTCGTTGAACCTGCGGCCGCGTTCGCTCGAGATC 831

Qy        268 IleAlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsnAlaLeu 287
              ::::::||||||:::|||   |||   :::   ||||||            |||   :::
Db        832 GTGTCTTGCCCGTCGTGCGGTCGCGCGCAAGTCGACGTCTACACCCTGGCCAACGAGGTA 891

Qy        288 GluGlnArgLeuGluAspIleIleThrProMetAspValSerIleIleGlyCysValVal 307
                  |||:::   :::      |||:::   |||:::::::::||||||||||||
Db        892 ACCGCCGGCCTGGATGGTCTCGATGTGCCGTTGCGGGTGGCCGTGATGGGGTGTGTCGTC 951

Qy        308 AsnGlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyGlyAsnLysLysSer 327
              ||||||||||||||||||      :::   |||||||||      ||||||   |||
```

```
Db        952 AATGGTCCGGGTGAAGCACGTGAGGCCGACCTGGGCGTGGCGTCCGGCAACGGCAAAGGT 1011

Qy        328 GlyLeuTyrGluAspGlyValArgLysAspArgLeuAspAsnAsnAspMetIleAspGln 347
                 ::::::      |||                :::                  :::::::::
Db       1012 CAGATCTTTGTACGGGGCGAAGTGATCAAGACCGTGCCCGAAGCACAGATCGTCGAGACG 1071

Qy        348 LeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAspGluAlaArgArgIleAspVal 367
                 |||                                :::::::||||||   |||:::
Db       1072 CTG----------------------------ATCGAGGAGGCGATGCGGCTGGCCGCC 1101

Qy        368 GlnGlnValGluLys 372
                 :::      |||:::
Db       1102 GAAATGGGCGAGCAA 1116

RESULT 12
US-09-921-992-1
; Sequence 1, Application US/09921992
; Patent No. US20020069426A1
; GENERAL INFORMATION:
;  APPLICANT: Boronat, Albert;
;  APPLICANT:  Campos, Narciso;
;  APPLICANT:  Rodriguez-Concepcion, Manuel;
;  APPLICANT:  Rohmer, Michel;
;  APPLICANT:  Seeman, Myriam;
;  APPLICANT:  Valentin, Henry E.;
;  APPLICANT:  Venkatesh, Tyamagondlu V.;
;  APPLICANT:  Venkatramesh, Mylavarapu
;  TITLE OF INVENTION: Methyl-D-Erythritol Phosphate Pathway Genes
;  FILE REFERENCE: 16516.107/35-21(51897)US
;  CURRENT APPLICATION NUMBER: US/09/921,992
;  CURRENT FILING DATE:  2001-08-06
;  PRIOR APPLICATION NUMBER: US 60/223,483
;  PRIOR FILING DATE: 2000-08-07
;  NUMBER OF SEQ ID NOS: 85
; SEQ ID NO 1
;   LENGTH: 2520
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (154)..(2376)
US-09-921-992-1
```

Alignment Scores:

Pred. No.:	1.66e-45	Length:	2520
Score:	462.00	Matches:	105
Percent Similarity:	55.88%	Conservative:	47
Best Local Similarity:	38.60%	Mismatches:	86
Query Match:	24.76%	Indels:	34
DB:	9	Gaps:	5

US-09-921-992-78 (1-372) x US-09-921-992-1 (1-2520)

```
Qy          9 ArgArgLysSerThrArgIleTyrValGlyAsnValProIleGlyAspGlyAlaProIle 28
              |||||||||:::      :::  ||||||||||||   :::|||         ||||||
Db        403 AGAAGGAAGACTCGTACTGTTATGGTTGGAAATGTCGCCCTTGGAAGCGAACATCCGATA 462

Qy         29 AlaValGlnSerMetThrAsnThrArgThrThrAspValGluAlaThrValAsnGlnIle 48
                 :::|||:::||||||   :::   |||   |||:::      ||||||:::::::::
Db        463 AGGATTCAAACGATGACTACTTCGGATACAAAAGATATTACTGGAACTGTTGATGAGGTT 522

Qy         49 LysAlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMetAspAla 68
                    :::         ||||||||||||||||||||:::::|||
Db        523 ATGAGAATAGCGGATAAAGGAGCTGATATTGTAAGGATAACTGTTCAAGGGAAGAAAGAG 582

Qy         69 AlaGluAlaPhe--------------LysLeuIleLysGlnGlnValAsnValProLeu 83
              |||:::|||                 ||||||:::::          |||:::||||||
Db        583 GCGGATGCGTGCTTTGAAATAAAAGATAAACTCGTTCAGCTTAATTACAATATACCGCTG 642

Qy         84 ValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAsp 103
              |||||||||||||||||      :::||||||:::||||||||          |||
Db        643 GTTGCAGATATTCATTTTGCCCCTACTGTAGCCTTACGAGTCGCTGAATGC---TTTGAC 699
```

```
Qy        104 CysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMetValValAsp 123
              :::|||:::||||||||||||      :::                  :::|||
Db        700 AAGATCCGTGTCAACCCAGGAAATTTTGCGGACAGGCGGGCCCAGTTTGAGACGATAGAT 759

Qy        124 CysAlaArgAspLys----------------------------------------------- 128
                  |||:::
Db        760 TATACAGAAGATGAATATCAGAAAGAACTCCAGCATATCGAGCAGGTCTTCACTCCTTTG 819

Qy        129 ----------------------AsnIleProIleArgIleGlyValAsnAlaGlySerLeu 141
                                     :::|||||||||   |||   ||||||||
Db        820 GTTGAGAAATGCAAAAAGTACGGGAGAGCAATGCGTATTGGGACAAATCATGGAAGTCTT 879

Qy        142 GluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSerAla 161
                  :::       |||||||::  :::|||::  ::::::||||||||
Db        880 TCTGACCGTATCATGAGCTATTACGGGGAT---TCTCCCCGAGGAATGGTTGAATCTGCG 936

Qy        162 MetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerValLys 181
                                  :::|||::::::      |||      |||:::|||
Db        937 TTTGAGTTTGCAAGAATATGTCGGAAATTAGACTATCACAACTTTGTTTTCTCAATGAAA 996

Qy        182 AlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIle------ 199
              |||||||:::      :::|   |||::::::||||||||||||||      ::::::
Db        997 GCGAGCAACCCAGTGATCATGGTCCAGGCGTACCGTTTACTTGTGGCTGAGATGTATGTT 1056

Qy        200 ---------AspGlnProLeuHisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGly 216
                      |||   |||||||||||||||:::||||||||||||||              |||
Db        1057 CATGGATGGGATTATCCTTTGCATTTGGGAGTTACTGAGGCAGGAGAAGGCGAAGATGGA 1116

Qy        217 AlaValLysSerAlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeu 236
              :::|||||||||||||||:::|||   ||||||   :::|||:::||||||||||:::
Db        1117 CGGATGAAATCTGCGATTGGAATTGGGACGCTTCTTCAGGACGGGCTCGGTGACACAATA 1176

Qy        237 ArgValSerLeuAlaAlaAspProValGluGluIle 248
              ||||||||||||      |||   |||||||||
Db        1177 AGAGTTTCACTGACGGAGCCACCAGAAGAGGAGATA 1212

RESULT 13
US-10-066-543-1184
; Sequence 1184, Application US/10066543
; Publication No. US20030087818A1
; GENERAL INFORMATION:
;  APPLICANT: Jiang, Yuqiu
;  APPLICANT:  Pyle, Ruth A.
;  APPLICANT:  Xu, Jiangchun
;  APPLICANT:  Indirias, Carol Yoseph
;  APPLICANT:  Lodes, Michael J.
;  APPLICANT:  Secrist, Heather
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Fanger, Gary R.
;  APPLICANT:  Smith, Carole L.
;  APPLICANT:  Durham, Margarita
;  APPLICANT:  Stolk, John A.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF COLON CANCER
;  FILE REFERENCE: 210121.563
;  CURRENT APPLICATION NUMBER: US/10/066,543
;  CURRENT FILING DATE:  2002-01-31
;  NUMBER OF SEQ ID NOS: 3417
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1184
;   LENGTH: 706
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 2, 645, 659
;   OTHER INFORMATION: n = A,T,C or G
US-10-066-543-1184

Alignment Scores:
Pred. No.:              4e-45       Length:        706
Score:                  452.00      Matches:       89
Percent Similarity:     73.29%      Conservative:  29
Best Local Similarity:  55.28%      Mismatches:    43
Query Match:            24.22%      Indels:        0
DB:                     14          Gaps:          0

US-09-921-992-78 (1-372) x US-10-066-543-1184 (1-706)

Qy         10 ArgLysSerThrArgIleTyrValGlyAsnValProIleGlyAspGlyAlaProIleAla 29
              ::::::|||      |||  :::|||   :::   :::|||   |||   |||:::
Db        218 AAAGAGAGTCGAGAAATACAAATCGGGATGTTAAGATTGGGAAAAGGAAATCCCGTTGTG 277

Qy         30 ValGlnSerMetThrAsnThrArgThrThrAspValGluAlaThrValAsnGlnIleLys 49
              :::|||||||||   ||||||   |||   ||||||||||||:::|||   ||||||
Db        278 ATTCAATCGATGATTAACACGGAAACTGCGGATGTAGAAGCGAGTGTCAGGCAAATTTTG 337

Qy         50 AlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMetAspAlaAla 69
              ||||||      |||   ::::::||||||:::::::::   |||   :::||||||
Db        338 GATTTGGAAGAAGCCGGCTGTGAATTGGTTCGTATGACGATCAATACAAAAGAGGCGGCT 397

Qy         70 GluAlaPheLysLeuIleLysGlnGlnValAsnValProLeuValAlaAspIleHisPhe 89
              |||      ||||||::::::|||::::::||||||||||||||||||||||||
Db        398 ATGGCAATTCCCGCGATTAAAAAGAGAGTTCATATTCCTTTGGTAGCGGATATTCATTTT 457

Qy         90 AspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArgIleAsnPro 109
              ||||||||||:::||||||         |||   |||:::|||   ||||||||||||||
Db        458 GATTATCGACTTGCCTTATTGGCGATAGAACAGGGAATTGACAAATTACGGATTAATCCC 517

Qy        110 GlyAsnIleGlyAsnGluGluArgIleArgMetValValAspCysAlaArgAspLysAsn 129
              ||||||||||||:::|||||||   |||||||:::|||   :::   |||:::   |||
Db        518 GGAAATATCGGTTCTGAAGAAAATATTCGTTTGGTTGCGGAAGCGGCAAAAAGGAAGAGA 577

Qy        130 IleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGlnGluLysTyr 149
              ||||||||||||||||||||||||||||||||||||||||||||||   :::   |||||||||
Db        578 ATTCCGATTCGTATTGGAGTCAATGCAGGTTCTTTGGAAAAACATATCTTGGAAAAATAT 637

Qy        150 GlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeuAsp 169
              |||      |||      |||   ::::::||||||||   ||||||      |||:::
Db        638 GGAGCGGNAACGGCAGATGCTNTGGTCAAAAGCGCTATGTATCATGTAAAACTGTTGGAA 697

Qy        170 Arg 170
              :::
Db        698 CAA 700

RESULT 14
US-09-921-992-22
; Sequence 22, Application US/09921992
; Patent No. US20020069426A1
; GENERAL INFORMATION:
;  APPLICANT: Boronat, Albert;
;  APPLICANT:  Campos, Narciso;
;  APPLICANT:  Rodriguez-Concepcion, Manuel;
;  APPLICANT:  Rohmer, Michel;
;  APPLICANT:  Seeman, Myriam;
;  APPLICANT:  Valentin, Henry E.;
;  APPLICANT:  Venkatesh, Tyamagondlu V.;
;  APPLICANT:  Venkatramesh, Mylavarapu
;  TITLE OF INVENTION: Methyl-D-Erythritol Phosphate Pathway Genes
;  FILE REFERENCE: 16516.107/35-21(51897)US
;  CURRENT APPLICATION NUMBER: US/09/921,992
;  CURRENT FILING DATE:  2001-08-06
;  PRIOR APPLICATION NUMBER: US 60/223,483
;  PRIOR FILING DATE: 2000-08-07
;  NUMBER OF SEQ ID NOS: 85
; SEQ ID NO 22
;   LENGTH: 670
;   TYPE: DNA
;   ORGANISM: Zea mays
US-09-921-992-22

Alignment Scores:
Pred. No.:              1.84e-31    Length:        670
Score:                  339.00      Matches:       81
Percent Similarity:     59.81%      Conservative:  44
Best Local Similarity:  38.76%      Mismatches:    60
```

```
Query Match:          18.17%      Indels:        24
DB:                   9           Gaps:          5

US-09-921-992-78 (1-372) x US-09-921-992-22 (1-670)

Qy        133 ArgIleGlyValAsnAlaGlySerLeuGlu-LysAspLeuGlnGluLysTyrGlyGluPr 152
              ||||||||||||||||  |||||||||    |||:::           |||
Db         15 AGGATCGGCGTTAACCATGGTTCTCTCTCCGAAAGAATGC------TTTTACCTACTTTT 68

Qy        152 oThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeuAspArgLeuAs 172
              ||||||:::   :::::::::||||||:::      :::         :::   |||::
Db         69 TACCCCCGAGGGCATGGTGCAATCGGCCCTGGAATTCATCAAAATTTGTGAGTCCTTAGA 128

Qy        172 nPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAlaValGluSerTy 192
              :|||          ||||||:::||||||||||   |||   :::   :::   :::||
Db        129 TTTCCGCAACCTAGTCGTTTCCATGAAAGCGTCCCGGGTACCGGTAATGTTGGCCGCCTA 188

Qy        192 rArgLeuLeuAlaLysGlnIleAspGln--------------ProLeuHisLeuGlyIl 207
              |||||||:::   |||::::::|||:::              ||||||||||||||::
Db        189 TCGCCTCATGGTGAAACGTATGGACGAGTTGGGCATGGATTATCCCCTCCATCTAGGGGT 248

Qy        207 eThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeuLe 227
              :||||||||||||          |||   :::||||||     |||:::      ||
Db        249 TACCGAAGCCGGGGATGGGGAATATGGCCGCATTAAATCCACCGCTGGCATTGCCACCCT 308

Qy        227 uLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGluGl 247
              ||||::::::|||||||||||||||:::||||||||||||          |||   :::||
Db        309 TTTAGCTGATGGCATTGGCGATACTATCCGGGTATCCCTCACCGAAGCCCCCGAAAAAGA 368

Qy        247 uIleLysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsnPh 267
              ||||   |||   :::   |||||:::::|||  :::|||          :::   ::
Db        369 AATTCCCGTTTGCTACAGCATTCTCCAGGCGCTGGGTTTGCGGAAAACCATGGTGGAATA 428

Qy        267 eIleAlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsnAlaLe 287
              ::::||||||||||:::|||   |||      |||:::                ||
Db        429 TGTGGCCTGTCCTTCCTGTGGCCGCACGTTGTTCAAC--------------------TT 467

Qy        287 uGluGlnArgLeuGluAspIle-----------------IleThrProMetAsp-ValS 301
              ||||      |||   ::::::                :::|||   :::|||    :
Db        468 GGAAGACGTGTTACATGAAGTCCGAGATGCCACTAAACATCTAACGGGTTTAGACTTTCG 527

Qy        301 erIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyValT 321
              :::::::::||||||:::||||||||||||||||          :::      |||
Db        528 CCGTCATGGGCTGTATTGTCAATGGCCCCGGGGCAATGGCCGATGCCGACTATGGCTATG 587

Qy        321 hrGlyGlyAsnLysLysSerGly 328
              |||      |||:::::::|||
Db        588 TGGGT------AAACAAGCCGGT 604

RESULT 15
US-09-921-992-23
; Sequence 23, Application US/09921992
; Patent No. US20020069426A1
; GENERAL INFORMATION:
;  APPLICANT: Boronat, Albert;
;  APPLICANT:  Campos, Narciso;
;  APPLICANT:  Rodriguez-Concepcion, Manuel;
;  APPLICANT:  Rohmer, Michel;
;  APPLICANT:  Seeman, Myriam;
;  APPLICANT:  Valentin, Henry E.;
;  APPLICANT:  Venkatesh, Tyamagondlu V.;
;  APPLICANT:  Venkatramesh, Mylavarapu
;  TITLE OF INVENTION: Methyl-D-Erythritol Phosphate Pathway Genes
;  FILE REFERENCE: 16516.107/35-21(51897)US
;  CURRENT APPLICATION NUMBER: US/09/921,992
;  CURRENT FILING DATE:  2001-08-06
;  PRIOR APPLICATION NUMBER: US 60/223,483
;  PRIOR FILING DATE: 2000-08-07
;  NUMBER OF SEQ ID NOS: 85
; SEQ ID NO 23
;   LENGTH: 596
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1..596)
;   OTHER INFORMATION: unsure at all n locations
US-09-921-992-23

Alignment Scores:
Pred. No.:                5.82e-30      Length:          596
Score:                    326.00        Matches:         78
Percent Similarity:       58.62%        Conservative:    41
Best Local Similarity:    38.42%        Mismatches:      62
Query Match:              17.47%        Indels:          22
DB:                       9             Gaps:            4

US-09-921-992-78 (1-372) x US-09-921-992-23 (1-596)

Qy        133 ArgIleGlyValAsnAlaGlySerLeuGlu-LysAspLeuGlnGluLysTyrGlyGluPr 152
              ||||||||||||||||  |||||||||    |||:::           |||
Db         15 AGGATCGGCGTTAACCATGGTTCTCTCTCCGAAAGAATGC------TTTTACCTACTTTT 68

Qy        152 oThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeuAspArgLeuAs 172
              ||||||:::   :::::::::||||||:::      :::         :::   |||::
Db         69 TACCCCCGAGGGCATGGTGCAATCGGCCCTGGAATTCATCAAAATTTGTGAGTCCTTAGA 128

Qy        172 nPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAlaValGluSerTy 192
              :|||          ||||||:::||||||||||   |||   :::   :::   :::||
Db        129 TTTCCGCAACCTAGTCGTTTCCATGAAAGCGTCCCGGGTACCGGTAATGTTGGCCGCCTA 188

Qy        192 rArgLeuLeuAlaLysGlnIleAspGln--------------ProLeuHisLeuGlyIl 207
              |||||||:::   |||::::::|||:::              ||||||||||||||::
Db        189 TCGCCTCATGGTGAAACGTATGGACGAGTTGGGCATGGATTATCCCCTCCATCTAGGGGT 248

Qy        207 eThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeuLe 227
              :||||||||||||          |||   :::||||||     |||:::      ||
Db        249 TACCGAAGCCGGGGATGGGGAATATGGCCGCATTAAATCCACCGCTGGCATTGCCACCCT 308

Qy        227 uLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGluGl 247
              ||||::::::|||||||||||||||:::||||||||||||          |||   :::||
Db        309 TTTAGCTGATGGCATTGGCGATACTATCCGGGTATCCCTCACCGAAGCCCCCGAAAAAGA 368

Qy        247 uIleLysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsnPh 267
              ||||   |||   :::   |||||:::::::|||   :::|||          :::   ::
Db        369 AATTCCCGTTTGCTACAGCATTCTCCAGGCGCTGGGTTTGCGGAAAACCATGGTGGAATA 428

Qy        267 eIleAlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsnAlaLe 287
              ::::||||||||||:::|||   |||      |||:::                   ||
Db        429 TGTGGCCTGTCCTTCCTGTGGCCGCACGTTGTTCAAC--------------------TT 467

Qy        287 uGluGlnArgLeuGluAspIle-----------------IleThrProMetAsp-ValS 301
              ||||      |||   ::::::                :::|||   :::|||
Db        468 GGAAGACGTGTTACATGAAGTCCGAGATGCCACTAAACATCTAACGTGTTTAGACTTTCG 527

Qy        301 erIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyValT 321
              ::::::   |||:::||||||||||||||||          :::      |||
Db        528 NCGTCATGTGCTGTATTGTCAATGGCCCCGGTGCAATGGCCGATGCCGACTATGGCTATG 587

Qy        321 hrGly 322
              |||
Db        588 TGGGT 592


Search completed: August 15, 2003, 10:15:58
Job time : 27217 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 12, 2003, 09:54:00 ; Search time 41 Seconds
(without alignments)
872.555 Million cell updates/sec

Title: US-09-921-992-78
Perfect score: 1866
Sequence: 1 MHNQAPIQRRKSTRIYVGNV.....RAKASQLDEARRIDVQQVEK 372

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1866	100.0	372	1	S23058	gcpE protein - Esc
2	1866	100.0	372	2	A91051	hypothetical prote
3	1866	100.0	372	2	E85895	hypothetical prote
4	1831	98.1	372	2	AB0822	GcpE protein (prot
5	1622.5	87.0	375	2	AG0350	probable acetyltra
6	1621.5	86.9	376	2	F82283	gcpE protein VC075
7	1603.5	85.9	368	1	H64063	gcpE protein - Hae
8	1372.5	73.6	371	2	F83171	conserved hypothet
9	1282	68.7	368	2	E84963	gcpE protein [impo
10	1042.5	55.9	383	2	H87354	gcpE protein [impo
11	887	47.5	349	2	G97121	gcpE protein [impo
12	882.5	47.3	367	2	A83825	peptidoglycan acet
13	868	46.5	368	2	AI1254	probable peptidogl
14	859.5	46.1	377	2	C69955	peptidoglycan acet
15	828	44.4	357	2	F70433	GcpE protein - Aqu
16	815.5	43.7	384	2	T35407	conserved hypothet
17	794	42.6	359	1	A64598	protein E - Helico
18	790	42.3	359	2	E71914	hypothetical prote
19	751.5	40.3	387	2	F70886	probable gcpE prot
20	712.5	38.2	392	2	G87106	conserved hypothet
21	706	37.8	344	2	D72321	gcpE protein - The
22	690	37.0	357	2	A81339	gcpE protein homol
23	618.5	33.1	403	2	S77159	gcpE protein - Syn
24	613.5	32.9	408	2	AF2118	hypothetical prote
25	588	31.5	404	2	F71324	probable gcpE prot
26	543.5	29.1	429	2	D75526	gcpE protein - Dei
27	527	28.2	421	2	D81098	gcpE protein NMB13
28	527	28.2	421	2	H81843	conserved hypothet
29	509	27.3	437	2	AH3285	gcpE protein [impo
30	504.5	27.0	601	2	C81715	gcpE protein TC032
31	492	26.4	416	2	AB2911	peptidoglycan acet
32	492	26.4	440	2	H97685	hypothetical prote
33	487	26.1	602	2	E71562	hypothetical prote
34	470.5	25.2	613	2	C86537	GcpE protein [impo
35	470.5	25.2	613	2	E72087	gcpE protein CP038
36	459	24.6	417	2	B82542	conserved hypothet
37	126	6.8	533	2	S76101	hypothetical prote
38	114	6.1	524	2	H69099	conserved hypothet
39	112.5	6.0	686	2	H90226	minichromosome mai
40	111	5.9	486	2	A75045	2-isopropylmalate
41	108	5.8	690	2	E72337	translation initia
42	106.5	5.7	284	2	D86701	mathionine aminope
43	106.5	5.7	531	2	S52294	2-isopropylmalate
44	106.5	5.7	531	2	AH2410	2-isopropylmalate
45	106	5.7	748	2	A64319	carbon-monoxide de

ALIGNMENTS

RESULT 1
S23058
gcpE protein - Escherichia coli (strain K-12)
C;Species: Escherichia coli
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 01-Mar-2002
C;Accession: S23058; B65028
R;Baker, J.; Franklin, D.B.; Parker, J.
FEMS Microbiol. Lett. 94, 175-180, 1992
A;Title: Sequence and characterization of the gcpE gene of Escherichia coli.
A;Reference number: S23057
A;Accession: S23058
A;Molecule type: DNA
A;Residues: 1-372 <BAK>
A;Cross-references: EMBL:X64451; NID:g41540; PID:g41542
A;Experimental source: strain K-12
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; .A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: B65028
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-372 <BLAT>
A;Cross-references: GB:AE000338; GB:U00096; NID:g1788862; PIDN:AAC75568.1; PID:g17888
A;Experimental source: strain K-12, substrain MG1655
C;Genetics:
A;Gene: gcpE
C;Superfamily: gcpE protein

Query Match 100.0%; Score 1866; DB 1; Length 372;
Best Local Similarity 100.0%; Pred. No. 1.3e-121;
Matches 372; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MHNQAPIQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MHNQAPIQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVR 60

Qy         61 VSVPTMDAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRM 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VSVPTMDAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRM 120

Qy        121 VVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSV 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 VVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSV 180

Qy        181 KASDVFLAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 KASDVFLAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSL 240

Qy        241 AADPVEEIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        241 AADPVEEIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDV 300

Qy        301 SIIGCVVNGPGEALVSTLGVTGGNKKSGLYEDGVRKDRLDNNDMIDQLEARIRAKASQLD 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SIIGCVVNGPGEALVSTLGVTGGNKKSGLYEDGVRKDRLDNNDMIDQLEARIRAKASQLD 360

Qy        361 EARRIDVQQVEK 372
              ||||||||||||
Db        361 EARRIDVQQVEK 372


RESULT 2
A91051
hypothetical protein ECs3377 [imported] - Escherichia coli (strain O157:H7, substrain RI
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C;Accession: A91051
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: A91051
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-372 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB36800.1; PID:g13362847; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs3377
C;Superfamily: gcpE protein

  Query Match             100.0%;  Score 1866;  DB 2;  Length 372;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-121;
  Matches  372;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MHNQAPIQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MHNQAPIQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVR 60

Qy         61 VSVPTMDAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRM 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VSVPTMDAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRM 120

Qy        121 VVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSV 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 VVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSV 180

Qy        181 KASDVFLAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 KASDVFLAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSL 240

Qy        241 AADPVEEIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AADPVEEIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDV 300

Qy        301 SIIGCVVNGPGEALVSTLGVTGGNKKSGLYEDGVRKDRLDNNDMIDQLEARIRAKASQLD 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SIIGCVVNGPGEALVSTLGVTGGNKKSGLYEDGVRKDRLDNNDMIDQLEARIRAKASQLD 360

Qy        361 EARRIDVQQVEK 372
              ||||||||||||
Db        361 EARRIDVQQVEK 372


RESULT 3
E85895
hypothetical protein gcpE [imported] - Escherichia coli (strain O157:H7, substrain EDL93
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: E85895
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: E85895
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-372 <STO>
A;Cross-references: GB:AE005174; NID:g12516909; PIDN:AAG57625.1; GSPDB:GN00145; UWGP:
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: gcpE
C;Superfamily: gcpE protein

  Query Match             100.0%;  Score 1866;  DB 2;  Length 372;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-121;
  Matches  372;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MHNQAPIQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MHNQAPIQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVR 60

Qy         61 VSVPTMDAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRM 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VSVPTMDAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRM 120

Qy        121 VVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSV 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 VVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSV 180

Qy        181 KASDVFLAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 KASDVFLAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSL 240

Qy        241 AADPVEEIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AADPVEEIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDV 300

Qy        301 SIIGCVVNGPGEALVSTLGVTGGNKKSGLYEDGVRKDRLDNNDMIDQLEARIRAKASQLD 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SIIGCVVNGPGEALVSTLGVTGGNKKSGLYEDGVRKDRLDNNDMIDQLEARIRAKASQLD 360

Qy        361 EARRIDVQQVEK 372
              ||||||||||||
Db        361 EARRIDVQQVEK 372


RESULT 4
AB0822
GcpE protein (protein E) [imported] - Salmonella enterica subsp. enterica serovar Typ
C;Species: Salmonella enterica subsp. enterica serovar Typhi
A;Note: this species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C;Accession: AB0822
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr
, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica se
A;Reference number: AB0502; MUID:21534947; PMID:11677608
A;Accession: AB0822
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-372 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD02726.1; PID:g16503738; GSPDB:GN00176
C;Genetics:
A;Gene: STY2768
C;Superfamily: gcpE protein

  Query Match             98.1%;  Score 1831;  DB 2;  Length 372;
  Best Local Similarity   98.1%;  Pred. No. 3.5e-119;
  Matches  365;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;
```

```
Qy        1 MHNQAPIQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVR 60
Db        1 MHNQAPIQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVR 60

Qy       61 VSVPTMDAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRM 120
Db       61 VSVPTMDAAEAFKLIKQQVSVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRM 120

Qy      121 VVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSV 180
Db      121 VVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSV 180

Qy      181 KASDVFLAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSL 240
Db      181 KASDVFLAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSL 240

Qy      241 AADPVEEIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDV 300
Db      241 AADPVEEIKVGFDILKSLRIRARGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDV 300

Qy      301 SIIGCVVNGPGEALVSTLGVTGGNKKSGLYEDGVRKDRLDNNDMIDQLEARIRAKASQLD 360
Db      301 SIIGCVVNGPGEALVSTLGVTGGNKKSGLYEDGVRKDRLDNDDMIAQLESRIRAKVSQLD 360

Qy      361 EARRIDVQQVEK 372
Db      361 EARRIDVLQVEK 372
```

RESULT 5
AG0350
probable acetyltransferase aarC [imported] - Yersinia pestis (strain CO92)
C;Species: Yersinia pestis
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Nov-2001
C;Accession: AG0350
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
Nature 413, 523-527, 2001
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A;Reference number: AB0001; MUID:21470413; PMID:11586360
A;Accession: AG0350
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-375 <KUR>
A;Cross-references: GB:AL590842; PIDN:CAC92130.1; PID:g15980846; GSPDB:GN00175
C;Genetics:
A;Gene: aarC
C;Superfamily: gcpE protein

```
  Query Match             87.0%;  Score 1622.5;  DB 2;  Length 375;
  Best Local Similarity   86.8%;  Pred. No. 9.1e-105;
  Matches  323;  Conservative   19;  Mismatches   29;  Indels    1;  Gaps    1;

Qy        1 MHNQAPIQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVR 60
Db        1 MHNGSPIIRRKSTRIYVGKVPIGDGAPIAVQSMTNTKTTDVDATVAQIKALERVGVDIVR 60

Qy       61 VSVPTMDAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRM 120
Db       61 VSIPTMDAAEAFKLIKQQSTVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNESRIRE 120

Qy      121 VVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSV 180
Db      121 VVACARDYNIPIRIGINGGSLEKDIQEKYGEPTPEALLESAMRHVDILDRLNFDQFKVSV 180

Qy      181 KASDVFLAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSL 240
Db      181 KASDVFLAVNSYRLLAKQINNPLHLGITEAGGARSGSVKSAIGLGLLLSEGIGDTLRISL 240

Qy      241 AADPVEEIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDV 300
Db      241 AADPVEEVKVGFDILKSLRIRARGINFIACPTCSRQEFDVIGTVNALEQRLEDLITPMDV 300

Qy      301 SIIGCVVNGPGEALVSTLGVTGGNKKSGLYEDGVR-KDRLDNNDMIDQLEARIRAKASQL 359
Db      301 SIIGCVVNGPGEALVSTIGVTGARNHSGFYEDGVRQKERFDNKAMIDQLEAKIRAKASIL 360

Qy      360 DEARRIDVQQVE 371
Db      361 DANNRIVINQLD 372
```

RESULT 6
F82283
gcpE protein VC0759 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
C;Species: Vibrio cholerae
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C;Accession: F82283
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A;Reference number: A82035; MUID:20406833; PMID:10952301
A;Accession: F82283
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-376 <HEI>
A;Cross-references: GB:AE004161; GB:AE003852; NID:g9655200; PIDN:AAF93924.1; GSPDB:GN
A;Experimental source: serogroup O1; strain N16961; biotype El Tor
C;Genetics:
A;Gene: VC0759
A;Map position: 1
C;Superfamily: gcpE protein

```
  Query Match             86.9%;  Score 1621.5;  DB 2;  Length 376;
  Best Local Similarity   85.3%;  Pred. No. 1.1e-104;
  Matches  318;  Conservative   33;  Mismatches   21;  Indels    1;  Gaps    1;

Qy        1 MHNQAPIQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVR 60
Db        3 MQHESPIKRRPSTRIYVGDVPIGDGAPIAVQSMTNTRTTDVAATVAQIKSLEKVGADIVR 62

Qy       61 VSVPTMDAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRM 120
Db       63 VSVPTMEAAEAFKLIKQQVSVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEARIRS 122

Qy      121 VVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSV 180
Db      123 VVDCARDKGIPIRIGVNGGSLEKDLQLKYGEPTPEALVESAMRHVDILDRLNFDQFKVSV 182

Qy      181 KASDVFLAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSL 240
Db      183 KASDVFLAVDSYRLLAKKIDQPLHLGITEAGGARAGSVKSAVGLGMLLAEGIGDTLRISL 242

Qy      241 AADPVEEIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDV 300
Db      243 AADPVEEIKVGFDILKSLRIRSRGINFIACPSCSRQEFDVIGTVNALEQRLEDVLTPMDV 302

Qy      301 SIIGCVVNGPGEALVSTLGVTGGNKKSGLYEDGVR-KDRLDNNDMIDQLEARIRAKASQL 359
Db      303 SIIGCVVNGPGEAEVSHLGLAGSNKKSAFYEDGVRQKERFDNDDLVAQLEAKIRAKAARL 362

Qy      360 DEARRIDVQQVEK 372
Db      363 DEKNRIDIKHVEQ 375
```

RESULT 7
H64063
gcpE protein - Haemophilus influenzae (strain Rd KW20)
C;Species: Haemophilus influenzae
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 28-Jul-2000
C;Accession: H64063
R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage

; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J
, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A;Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
A;Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A;Reference number: A64000; MUID:95350630; PMID:7542800
A;Accession: H64063
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-368 <TIGR>
A;Cross-references: GB:U32721; GB:L42023; NID:g1573334; PIDN:AAC22026.1; PID:g1573337; T
C;Genetics:
A;Gene: gcpE
C;Superfamily: gcpE protein

Query Match 85.9%; Score 1603.5; DB 1; Length 368;
Best Local Similarity 88.2%; Pred. No. 1.8e-103;
Matches 320; Conservative 20; Mismatches 22; Indels 1; Gaps 1;

```
Qy          4 QAPIQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVRVSV 63
              |  |:||:||:||||||||||  ||||||||||||||||||| |||:||||||||||||||
Db          5 QPTIKRRESTKIYVGNVPIGGDAPIAVQSMTNTRTTDVEATVAQIKSLERVGADIVRVSV 64

Qy         64 PTMDAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRMVVD 123
              |||||||||| ||||||||||||||||||||||||||||||||||||||| |:|:| |||
Db         65 PTMDAAEAFKQIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGREDRVRAVVD 124

Qy        124 CARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSVKAS 183
              ||||||||||||||||||||||||||||||:||||||:|||: ||||||||||||||||
Db        125 CARDKNIPIRIGVNAGSLEKDLQEKYGEPTPEALLESALRHVEILDRLNFDQFKVSVKAS 184

Qy        184 DVFLAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSLAAD 243
              |||||||||||||| |||||||||||||||:||||||:|||:||:|||||||||||||||
Db        185 DVFLAVESYRLLAKAIKQPLHLGITEAGGARAGAVKSAVGLGMLLAEGIGDTLRVSLAAD 244

Qy        244 PVEEIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDVSII 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        245 PVEEIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDVSII 304

Qy        304 GCVVNGPGEALVSTLGVTGGNKKSGLYEDGVR-KDRLDNNDMIDQLEARIRAKASQLDEA 362
              |||||||||||||| |||||||||| | || | |:| || |:::||||:|||| :: |
Db        305 GCVVNGPGEALVSDLGVTGGNKKSGYYLDGERQKERFDNEDIVNQLEAKIRAKVARQDPK 364

Qy        363 RRI 365
              ||
Db        365 NRI 367
```

RESULT 8
F83171
conserved hypothetical protein PA3803 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: F83171
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: F83171
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-371 <STO>
A;Cross-references: GB:AE004798; GB:AE004091; NID:g9949963; PIDN:AAG07190.1; GSPDB:GN001
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA3803
C;Superfamily: gcpE protein

Query Match 73.6%; Score 1372.5; DB 2; Length 371;
Best Local Similarity 73.8%; Pred. No. 1.7e-87;
Matches 267; Conservative 46; Mismatches 48; Indels 1; Gaps 1;

```
Qy          1 MHNQAPIQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVR 60
              :|: :|| |||| :|:|||||:|  |||||||||| | || ||| ||: ||  ||||||
Db          3 IHSASPIIRRKSRKIWVGNVPVGGDAPIAVQSMTNTETCDVAATVAQIRRLEDAGADIVR 62

Qy         61 VSVPTMDAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRM 120
              |||| |||||||  ||||||||||||||||||||||:||| |||||||||||| |:|::
Db         63 VSVPDMDAAEAFGKIKQQVNVPLVADIHFDYRIALRVAELGVDCLRINPGNIGREDRVKA 122

Qy        121 VVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSV 180
              ||| ||::|||||||||||||||||:|||||||:|||||||||||||||:|:|  |||||
Db        123 VVDAARERNIPIRIGVNAGSLEKDLQKKYGEPTPEALLESAMRHVDHLDKLDFQNFKVSV 182

Qy        181 KASDVFLAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSL 240
              ||||||:|| :|||||:||:|||||||||||| ||| ||||:|||:||:|||||:|:||
Db        183 KASDVFMAVAAYRLLARQIEQPLHLGITEAGGLRSGTVKSAVGLGMLLAEGIGDTIRISL 242

Qy        241 AADPVEEIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDV 300
              |||||||||||||||||| :|||||||||||:|||| |||: |:| || ||||:: ||||
Db        243 AADPVEEIKVGFDILKSLHLRSRGINFIACPSCSRQNFDVVKTMNELEGRLEDLLVPMDV 302

Qy        301 SIIGCVVNGPGEALVSTLGVTGGNKKSGLYEDGVRKDRLDNNDMIDQLEARIRAKASQLD 360
              ::|||||||||||  : :|:|||     :| ||    :| |::::|:|| || ||::
Db        303 AVIGCVVNGPGEAKEAHVGLTGGTPNL-VYIDGKPSQKLTNDNLVDELERLIRQKAAEKA 361

Qy        361 EA 362
              ||
Db        362 EA 363
```

RESULT 9
E84963
gcpE protein [imported] - Buchnera sp. (strain APS)
C;Species: Buchnera sp.
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
C;Accession: E84963
R;Shigenobu, S.; Watanabe, H.; Hattori, M.; Sakaki, Y.; Ishikawa, H.
Nature 407, 81-86, 2000
A;Title: Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp
A;Reference number: A84930; MUID:20445173; PMID:10993077
A;Accession: E84963
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-368 <STO>
A;Cross-references: GB:AP000398; GSPDB:GN00144
A;Experimental source: strain APS
C;Genetics:
A;Gene: gcpE; BU287
C;Superfamily: gcpE protein

Query Match 68.7%; Score 1282; DB 2; Length 368;
Best Local Similarity 66.1%; Pred. No. 3e-81;
Matches 242; Conservative 65; Mismatches 57; Indels 2; Gaps 2;

```
Qy          1 MHNQAP-IQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIV 59
              |:|:   | |||| ||||| | ||: |||:||||||||||::  |:|||  |::|| |||
Db          1 MNNKCKIINRRKSDRIYVGKVAIGNNAPISVQSMTNTRTTNISETINQILELQKVGVDIV 60

Qy         60 RVSVPTMDAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIR 119
              |:|:| : |||:|| ||:| ||||:||||||||:||:  :|| ||||||||||||: |:
Db         61 RISIPNLKAAESFKEIKKQTNVPLIADIHFDYRLALQAIKYGADCLRINPGNIGNKRRVS 120

Qy        120 MVVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVS 179
              ::  |:|:|||||||||||||||: :||  ||| ||:||||||::: | |||:|||||
Db        121 EIISYAKDENIPIRIGVNAGSLEKDILKKYKIPTPDALVESAMRHIEYFDALNFNQFKVS 180

Qy        180 VKASDVFLAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVS 239
              |||||||||:||||:| |:| ||||:||||:|| |:| |||:||: ||| |||||||:|||
Db        181 VKASDVFLAIESYRMLGKKITQPLHIGITESGGLRNGTVKSSIGIALLLLEGIGDTIRVS 240

Qy        240 LAADPVEEIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMD 299
              ||| | ||:|||:|||| | :|:||||||||||||||||||| ||| ||: |||| ||:|
Db        241 LAAHPTEEVKVGYDILKVLSLRARGINFIACPTCSRQEFDVINTVNQLEKNLEDISTPID 300
```

```
Qy        300 VSIIGCVVNGPGEALVSTLGVTGGNKKSGLYEDGVR-KDRLDNNDMIDQLEARIRAKASQ 358
              |||||||||| ||: ::|||: | :|||  |||||| |::: | ::|:::| :|| |  :
Db        301 VSIIGCVVNGIGESKIATLGLAGSHKKSAFYEDGVRQKEKIKNEEIIEKMEIKIRKKIDK 360

Qy        359 LDEARR 364
              |: :::
Db        361 LNNSKK 366


RESULT 10
H87354
gcpE protein [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C;Accession: H87354
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: H87354
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-383 <STO>
A;Cross-references: GB:AE005673; NID:g13422106; PIDN:AAK22836.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC0851
C;Superfamily: gcpE protein

  Query Match             55.9%;  Score 1042.5;  DB 2;  Length 383;
  Best Local Similarity   57.6%;  Pred. No. 1.1e-64;
  Matches  208;  Conservative   57;  Mismatches   95;  Indels    1;  Gaps    1;

Qy          7 IQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVRVSVPTM 66
              | ||:| :| ||:| :|  |||:|||||| |:|  ||: ||: ||  |||||||| | :
Db         14 ITRRQSRKIRVGSVEVGGDAPISVQSMTNTLTSDAAATLEQIRQLEEAGADIVRVSCPDV 73

Qy         67 DAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRMVVDCAR 126
              ::  ||| | ::  |||||||| |:  :: |: |  |||||||||: :|:| |:  ||
Db         74 ESTAAFKTIAREAKVPLVADIHFHYKRGIEAAQAGAACLRINPGNIGSPDRVRDVIQAAR 133

Qy        127 DKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSVKASDVF 186
              |    :|||||||||::| |||||| | |::|||: |   |   :| :||:|||||| |
Db        134 DHGCSMRIGVNAGSLERELLEKYGEPCPDAMVESALNHARILQDHDFHEFKISVKASDPF 193

Qy        187 LAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSLAADPVE 246
              : | :|  ||: || |||||:|||  |:| ||||||:| :|  |||||:|||||||||
Db        194 MTVAAYYQLAEAIDCPLHLGVTEAGATRTGTVKSAIGIGAMLWAGIGDTIRVSLAADPVE 253

Qy        247 EIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDVSIIGCV 306
              |||||||||||| :| ||:| ||||:|:|| |:|| || |||:||  | ||| :||||||
Db        254 EIKVGFDILKSLGLRHRGVNIIACPSCARQGFNVIKTVEALEERLAHISTPMSLSIIGCV 313

Qy        307 VNGPGEALVSTLGVTGGNKKSGL-YEDGVRKDRLDNNDMIDQLEARIRAKASQLDEARRI 365
              ||||||||:: :| |||   :|: |  |    :  | ||| :   :  ||::: |:
Db        314 VNGPGEALMTDIGFTGGGAGAGMVYMAGKPDHKQSNEGMIDHIVDLVEKKAAEIQAAKAQ 373

Qy        366 D 366
              |
Db        374 D 374


RESULT 11
G97121
gcpE protein [imported] - Clostridium acetobutylicum
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001
C;Accession: G97121
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: G97121
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-349 <KUR>
A;Cross-references: GB:AE001437; PIDN:AAK79762.1; PID:g15024769; GSPDB:GN00168
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC1797
C;Superfamily: gcpE protein

  Query Match             47.5%;  Score 887;  DB 2;  Length 349;
  Best Local Similarity   48.8%;  Pred. No. 5.6e-54;
  Matches  169;  Conservative   75;  Mismatches  102;  Indels    0;  Gaps    0;

Qy          7 IQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVRVSVPTM 66
              :|| :: :: ||:| :|  : : ||||||| | | : |: ||| ||  | |||| :||
Db          1 MQRIETRKVKVGSVYVGGDSRVTVQSMTNTDTRDSKKTIEQIKKLEIAGCDIVRCAVPDF 60

Qy         67 DAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRMVVDCAR 126
              ||| :   | : | :|:|||||||||:||:  : ||  |||||||||::||: :|   |:
Db         61 DAASSLSEITKNVKLPVVADIHFDYRLALEAIKNGVSALRINPGNIGSKERVELVAKSAK 120

Qy        127 DKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSVKASDVF 186
              :||||||||||:||||||:  ||      :||:|||: |  |: :||:   :|:|:|:|
Db        121 EKNIPIRIGVNSGSLEKDILNKYKRVCSEALVESALNHVKILEDVNFNDIVISIKSSNVQ 180

Qy        187 LAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSLAADPVE 246
              : ::||||::|::: |||||:||||    | :||:||:| ||||||||:||||  ||||
Db        181 MMIDSYRLISKEVNYPLHLGVTEAGTIWRGTIKSSIGIGTLLSEGIGDTIRVSLTGDPVE 240

Qy        247 EIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDVSIIGCV 306
              |:||| :|||:      |: ||:|||| |   |:|   | :|:|||     : |:::|||
Db        241 EVKVGREILKTFGYLKSGVEFISCPTCGRTSIDLIKIANEVEKRLEKTNKSIKVAVMGCV 300

Qy        307 VNGPGEALVSTLGVTGGNKKSGLYEDGVRKDRLDNNDMIDQLEARI 352
              |||||||  : :|: ||  :  ::: |    ::   :::|:|   |
Db        301 VNGPGEAREADIGIAGGKGEGLIFKKGEIIKKVKEENLVDELMREI 346


RESULT 12
A83825
peptidoglycan acetylation gcpE [imported] - Bacillus halodurans (strain C-125)
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C;Accession: A83825
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; H
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a
A;Reference number: A83650; MUID:20512582; PMID:11058132
A;Accession: A83825
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-367 <STO>
A;Cross-references: GB:AP001511; GB:BA000004; NID:g10173727; PIDN:BAB05120.1; GSPDB:G
A;Experimental source: strain C-125
C;Genetics:
A;Gene: gcpE
C;Superfamily: gcpE protein

  Query Match             47.3%;  Score 882.5;  DB 2;  Length 367;
  Best Local Similarity   48.8%;  Pred. No. 1.2e-53;
  Matches  176;  Conservative   64;  Mismatches  114;  Indels    7;  Gaps    1;

Qy          9 RRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVRVSVPTMDA 68
              | |:  : |||: ||    : ||||| |:| |||||| :|| ||  |  :|||: | | |
Db          7 RTKTRPVKVGNLTIGGNNEVVVQSMTTTKTHDVEATVAEIKRLEEAGCQVVRVACPDMRA 66

Qy         69 AEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRMVVDCARDK 128
              |||   ||:|:::|||  ||||||::|||  | | | :|||||||   ::  ||  |::|
Db         67 AEAIPAIKKQISIPLVVDIHFDYKLALKAIEGGADKIRINPGNIGKRHKVEAVVKAAKEK 126
```

```
Qy        129 NIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSVKASDVFLA 188
              ||||||||||||||| 8 :||| ||   ::|||: |:  |: |:|    ||:|||| ||
Db        127 GIPIRIGVNAGSLEKRLLDKYGYPTADGMVESALHHIKILEDLDFHDIIVSMKASDVNLA 186

Qy        189 VESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSLAADPVEEI 248
              :|:|   ||  | |||||||:|   :| |||| ||| :|: |||:|:||||:||||||:
Db        187 IEAYEKAAKAFDYPLHLGITESGTLFAGTVKSAAGLGAILNMGIGNTVRVSLSADPVEEV 246

Qy        249 KVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDVSIIGCVVN 308
              ||  ::|||  : |   |:|||| | | |:|   | :|: :  |  |: |:::|| ||
Db        247 KVARELLKSFGLASNAATLISCPTCGRIEIDLISIANEMEEYISKIRAPIKVAVLGCAVN 306

Qy        309 GPGEALVSTLGVTGGNKKSGLYEDGVRKDRLDNNDMIDQLEARIR-------AKASQLDE 361
              |||||  : :|: |  :  |:  |    ::    |:::|:  |         ||  :|::
Db        307 GPGEAREADIGIAGARGEGLLFRKGEIVRKVPEETMVEELKKEIDKIAEEYFAKQKELEK 366

Qy        362 A 362
              |
Db        367 A 367


RESULT 13
AI1254
probable peptidoglycan acetylation protein homolog lmo1441 [imported] - Listeria monocyt
C;Species: Listeria monocytogenes
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C;Accession: AI1254
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AI1254
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-368 <GLA>
A;Cross-references: GB:NC_003210; PIDN:CAC99519.1; PID:g16410870; GSPDB:GN00177
A;Experimental source: strain EGD-e
C;Genetics:
A;Gene: lmo1441
C;Superfamily: gcpE protein

  Query Match             46.5%;  Score 868;  DB 2;  Length 368;
  Best Local Similarity   49.2%;  Pred. No. 1.2e-52;
  Matches  178;  Conservative   67;  Mismatches  115;  Indels    2;  Gaps    2;

Qy         10 RKSTR-IYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVRVSVPTMDA 68
              |::|| : |||: ||   : :|||| |:| |||||| :|  ||  |  ||||: |   |
Db          7 RENTRPVQVGNLTIGGSEELTIQSMTTTKTHDVEATVAEIHRLEEAGCQIVRVACPDERA 66

Qy         69 AEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRMVVDCARDK 128
              | |   ||:::::|||||||||:|||  : ||| :|||||||  :|:  ||: |: |
Db         67 ANALSAIKKKIHIPLVADIHFDYRLALKAIDAGVDKIRINPGNIGRRDRVEKVVNAAKAK 126

Qy        129 NIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSVKASDVFLA 188
              |||||||||||||| 8 :||| ||   ::|||: |:  |: |:|    ||:|||| ||
Db        127 NIPIRIGVNAGSLEKKIQKYGYPTADGMVESALAHIKILEDLDFYDIIVSLKASDVNLA 186

Qy        189 VESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSLAADPVEEI 248
              :|:|   ::  : |||||||:|   :| :||| ||| :|| |||:||||||:|||||||
Db        187 IEAYDKASRAFNYPLHLGITESGTQFAGGIKSAAGLGAILSLGIGNTLRVSLSADPVEEI 246

Qy        249 KVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDVSIIGCVVN 308
              ||  ::|||  : |   |:|||| | | |:|   | :|  :  |  |: |:::|| ||
Db        247 KVAREVLKSFGLSSNAAMLISCPTCGRIEIDLIRIANEVENYIAKIEVPIKVAVLGCAVN 306

Qy        309 GPGEALVSTLGVTGGNKKSGLYEDGVRKDRLDNNDMIDQLEARIRAKASQLDEARRIDVQ 368
              |||||  : :|: | |  :  |:  |    ::    |||:|:  |   | :    ::||::
Db        307 GPGEAREADIGIAGSNGEGLLFRHGKIIRKVPEAIMIDELKKEIDILAEEF-FVKKIDLE 365

Qy        369 QV 370
              ::
Db        366 SL 367


RESULT 14
C69955
peptidoglycan acetylation homolog yqfY - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 21-Jul-2000
C;Accession: C69955
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanl
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: C69955
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-377 <KUN>
A;Cross-references: GB:Z99116; GB:AL009126; NID:g2634723; PIDN:CAB14437.1; PID:g26349
A;Experimental source: strain 168
C;Genetics:
A;Gene: yqfY
C;Superfamily: gcpE protein

  Query Match             46.1%;  Score 859.5;  DB 2;  Length 377;
  Best Local Similarity   47.3%;  Pred. No. 4.9e-52;
  Matches  172;  Conservative   66;  Mismatches  115;  Indels   11;  Gaps    1;

Qy          9 RRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVRVSVPTMDA 68
              | |:  : || : ||    : :|||| |:| |||||| :|  |   |  ||||: |   |
Db          9 RTKTRPVKVGPLTIGGNNEVVIQSMTTTKTHDVEATVAEINRLAEAGCQIVRVACPDERA 68

Qy         69 AEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRMVVDCARDK 128
              | |   ||:::::||| ||||||::|||  | | | :|||||||  |::  ||  |:||
Db         69 ANAIADIKKRISIPLVVDIHFDYKLALKAIEGGADKIRINPGNIGRREKVEAVVKAAKDK 128

Qy        129 NIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSVKASDVFLA 188
              |||||||||||||| : |||| ||   ::|||: |:  |: |:|    ||:|||| ||
Db        129 GIPIRIGVNAGSLEKRILEKYGYPTADGMVESALHHIKILEDLDFHDIIVSMKASDVNLA 188

Qy        189 VESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSLAADPVEEI 248
              :|:|   ||  | |||||||:|   :| |||| ||| :||:|||:|:|:||:||||||:
Db        189 IEAYEKAAKAFDYPLHLGITESGTLFAGTVKSAAGLGAILSKGIGNTMRISLSADPVEEV 248

Qy        249 KVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDVSIIGCVVN 308
              ||  ::|||  : |   |:|||| | | |:|   | :|: :  |  |: |:::|| ||
Db        249 KVARELLKSFGLASNAATLISCPTCGRIEIDLISIANEVEEYISKIKAPIKVAVLGCAVN 308

Qy        309 GPGEALVSTLGVTGGNKKSGLYEDGVRKDRLDNNDMIDQLE----------ARIRAKAS 357
              |||||  : :|: |   :  |:  |    ::    |:::|:           |:: |: :
Db        309 GPGEAREADIGIAGARGEGLLFRKGKIVRKVPEETMVEELKKEIDILAEEHYAKLEAEKA 368

Qy        358 QLDE 361
              :| |
Db        369 KLKE 372


RESULT 15
F70433
GcpE protein - Aquifex aeolicus
```

```
C;Species: Aquifex aeolicus
C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 20-Sep-1999
C;Accession: F70433
R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ov
V.
Nature 392, 353-358, 1998
A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A;Reference number: A70300; MUID:98196666; PMID:9537320
A;Accession: F70433
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-357 <AQF>
A;Cross-references: GB:AE000745; NID:g2983907; PIDN:AAC07467.1; PID:g2983917; GB:AE00065
A;Experimental source: strain VF5
C;Genetics:
A;Gene: gcpE
C;Superfamily: gcpE protein

  Query Match             44.4%;  Score 828;  DB 2;  Length 357;
  Best Local Similarity   47.6%;  Pred. No. 6.9e-50;
  Matches  165;  Conservative   73;  Mismatches  109;  Indels    0;  Gaps    0;

Qy          7 IQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVRVSVPTM 66
              ||:||: :| |||| ||  ||| |||||:|:| |||||:|||| |   | :||||:||
Db          2 IQKRKTRQIRVGNVKIGGDAPIVVQSMTSTKTHDVEATLNQIKRLYEAGCEIVRVAVPHK 61

Qy         67 DAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRMVVDCAR 126
              :  || : | ::  :|::|||||    |    | ||  :||||||| || :| :|: |:
Db         62 EDVEALEEIVKKSPMPVIADIHFAPSYAFLSMEKGVHGIRINPGNIGKEEIVREIVEEAK 121

Qy        127 DKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSVKASDVF 186
               : : :|||||:|||||||| |||| |: :|| |||:|  :  ::  |  :|||:| |||
Db        122 RRGVAVRIGVNSGSLEKDLLEKYGYPSAEALAESALRWSEKFEKWGFTNYKVSIKGSDVL 181

Qy        187 LAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSLAADPVE 246
               |  :  : |:: | |||:||||||    | :||::|:|:|| |||||:|||| |||
Db        182 QNVRANLIFAERTDVPLHIGITEAGMGTKGIIKSSVGIGILLYMGIGDTVRVSLTDDPVV 241

Qy        247 EIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDVSIIGCV 306
              |::  ::||||| :| ||:  :||||| | | |:   |  ::::|  : ||: |:::|||
Db        242 EVETAYEILKSLGLRRRGVEIVACPTCGRIEVDLPKVVKEVQEKLSGVKTPLKVAVMGCV 301

Qy        307 VNGPGEALVSTLGVTGGNKKSGLYEDGVRKDRLDNNDMIDQLEARIR 353
              ||  |||  : :|:  |   : |:: |    ::| ::|:|:|   |:
Db        302 VNAIGEAREADIGLACGRGFAWLFKHGKPIKKVDESEMVDELLKEIQ 348

Search completed: August 12, 2003, 09:56:50
Job time : 43 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 12, 2003, 09:53:55 ; Search time 23 Seconds
(without alignments)
760.606 Million cell updates/sec

Title: US-09-921-992-78
Perfect score: 1866
Sequence: 1 MHNQAPIQRRKSTRIYVGNV.....RAKASQLDEARRIDVQQVEK 372

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1866	100.0	372	1	ISPG_ECOLI	P27433 escherichia
2	1840	98.6	372	1	ISPG_SALTY	P58671 salmonella
3	1831	98.1	372	1	ISPG_SALTI	P58670 salmonella
4	1645.5	88.2	372	1	ISPG_VIBPA	Q87s16 vibrio para
5	1631.5	87.4	372	1	ISPG_VIBVU	Q8dez8 vibrio vuln
6	1622.5	87.0	365	1	ISPG_PROST	P72241 providencia
7	1622.5	87.0	375	1	ISPG_YERPE	P58672 yersinia pe
8	1621.5	86.9	376	1	ISPG_VIBCH	Q9ktx1 vibrio chol
9	1603.5	85.9	368	1	ISPG_HAEIN	P44667 haemophilus
10	1599.5	85.7	367	1	ISPG_PASMU	P57987 pasteurella
11	1372.5	73.6	371	1	ISPG_PSEAE	Q9hxj4 pseudomonas
12	1285.5	68.9	367	1	ISPG_BUCAP	Q8k9p4 buchnera ap
13	1282	68.7	368	1	ISPG_BUCAI	P57374 buchnera ap
14	1226.5	65.7	366	1	ISPG_WIGBR	Q8dly3 wiggleswort
15	1042.5	55.9	383	1	ISPG_CAUCR	Q9a9w0 caulobacter
16	906	48.6	349	1	ISPG_CLOPE	P58667 clostridium
17	902.5	48.4	355	1	ISPG_THETN	Q8ra30 thermoanaer
18	887	47.5	349	1	ISPG_CLOAB	Q97i56 clostridium
19	882.5	47.3	367	1	ISPG_BACHD	Q9kd18 bacillus ha
20	869.5	46.6	354	1	ISPG_FUSNN	Q8rg40 fusobacteri
21	868	46.5	368	1	ISPG_LISMO	P58668 listeria mo
22	859.5	46.1	377	1	ISPG_BACSU	P54482 bacillus su
23	828	44.4	357	1	ISPG_AQUAE	O67496 aquifex aeo
24	815.5	43.7	384	1	ISG1_STRCO	Q9x7w2 streptomyce
25	815	43.7	385	1	ISG2_STRCO	Q9kyr9 streptomyce
26	794	42.6	359	1	ISPG_HELPY	O25342 helicobacte
27	790	42.3	359	1	ISPG_HELPJ	Q9zll0 helicobacte
28	761	40.8	378	1	ISPG_CORGL	Q8np12 corynebacte
29	751.5	40.3	387	1	ISPG_MYCTU	O33350 mycobacteri
30	712.5	38.2	392	1	ISPG_MYCLE	Q9cbu5 mycobacteri
31	706	37.8	344	1	ISPG_THEMA	Q9wzz3 thermotoga
32	690	37.0	357	1	ISPG_CAMJE	Q9ppm1 campylobact
33	618.5	33.1	403	1	ISPG_SYNY3	P73672 synechocyst
34	613.5	32.9	408	1	ISPG_ANASP	P58666 anabaena sp
35	588	31.5	404	1	ISPG_TREPA	O83460 treponema p
36	543.5	29.1	429	1	ISPG_DEIRA	Q9rxc9 deinococcus
37	527	28.2	421	1	ISPG_NEIMA	Q9ju34 neisseria m
38	527	28.2	421	1	ISPG_NEIMB	Q9jz40 neisseria m
39	512	27.4	417	1	ISPG_RHILO	Q98fg0 rhizobium l
40	509	27.3	420	1	ISPG_BRUME	Q8yj17 brucella me
41	504.5	27.0	601	1	ISPG_CHLMU	Q9pky3 chlamydia m
42	493	26.4	417	1	ISPG_RHIME	Q92l19 rhizobium m
43	493	26.4	431	1	ISPG_RALSO	P58669 ralstonia s
44	492	26.4	416	1	ISPG_AGRT5	P58665 agrobacteri
45	487	26.1	602	1	ISPG_CHLTR	O84060 chlamydia t

ALIGNMENTS

RESULT 1
ISPG_ECOLI
ID ISPG_ECOLI STANDARD; PRT; 372 AA.
AC P27433; P76984; P76985;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase (GcpE protein)
DE (Protein E).
GN ISPG OR GCPE OR B2515 OR C3037 OR Z3778 OR ECS3377.
OS Escherichia coli,
OS Escherichia coli O6, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 217992, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=92394488; PubMed=1521767;
RA Baker J., Franklin D.B., Parker J.;
RT "Sequence and characterization of the gcpE gene of Escherichia coli.";
RL FEMS Microbiol. Lett. 73:175-180(1992).
RN [2]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX MEDLINE=21625081; PubMed=11752431;
RA Hecht S., Eisenreich W., Adam P., Amslinger S., Kis K., Bacher A.,
RA Arigoni D., Rohdich F.;
RT "Studies on the nonmevalonate pathway to terpenes: the role of the
RT GcpE (IspG) protein.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:14837-14842(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97349980; PubMed=9205837;
RA Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,
RA Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasundaram S.,
RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
RA Yamagata S., Horiuchi T.;
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli
RT -K12 genome corresponding to 50.0-68.8 min on the linkage map and
RT analysis of its sequence features.";
RL DNA Res. 4:91-113(1997).
RN [5]

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX   MEDLINE=22388234; PubMed=12471157;
RA   Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA   Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA   Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA   Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT   "Extensive mosaic structure revealed by the complete genome sequence
RT   of uropathogenic Escherichia coli.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
RN   [6]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX   MEDLINE=21074935; PubMed=11206551;
RA   Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA   Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA   Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA   Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA   Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA   Welch R.A., Blattner F.R.;
RT   "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL   Nature 409:529-533(2001).
RN   [7]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / RIMD 0509952;
RX   MEDLINE=21156231; PubMed=11258796;
RA   Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA   Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA   Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA   Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT   "Complete genome sequence of enterohemorrhagic Escherichia coli
RT   O157:H7 and genomic comparison with a laboratory strain K-12.";
RL   DNA Res. 8:11-22(2001).
RN   [8]
RP   PRELIMINARY SEQUENCE OF 343-372 FROM N.A.
RX   MEDLINE=85261421; PubMed=2991272;
RA   Freedman R., Gibson B., Donovan D., Biemann K., Eisenbeis S.J.,
RA   Parker J., Schimmel P.;
RT   "Primary structure of histidine-tRNA synthetase and characterization
RT   of hisS transcripts.";
RL   J. Biol. Chem. 260:10063-10068(1985).
RN   [9]
RP   PATHWAY.
RC   STRAIN=K12 / MC4100;
RX   MEDLINE=21099853; PubMed=11163766;
RA   Campos N., Rodriguez-Concepcion M., Seemann M., Rohmer M., Boronat A.;
RT   "Identification of gcpE as a novel gene of the 2-C-methyl-D-erythritol
RT   4-phosphate pathway for isoprenoid biosynthesis in Escherichia coli.";
RL   FEBS Lett. 488:170-173(2001).
RN   [10]
RP   PATHWAY.
RX   MEDLINE=21172855; PubMed=11274098;
RA   Altincicek B., Kollas A.-K., Sanderbrand S., Wiesner J., Hintz M.,
RA   Beck E., Jomaa H.;
RT   "GcpE is involved in the 2-C-methyl-D-erythritol 4-phosphate pathway
RT   of isoprenoid biosynthesis in Escherichia coli.";
RL   J. Bacteriol. 183:2411-2416(2001).
CC   -!- FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate
CC       (ME-2,4cPP) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate.
CC   -!- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; sixth step.
CC   -!- SIMILARITY: BELONGS TO THE ISPG FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X64451; CAA45783.1; -.
DR   EMBL; AY033515; AAK53460.1; -.
DR   EMBL; AE000338; AAC75568.1; -.
DR   EMBL; D90880; BAA16402.1; -.
DR   EMBL; D90881; BAA20919.1; -.
DR   EMBL; AE016764; AAN81487.1; -.
DR   EMBL; AE005481; AAG57625.1; -.
DR   EMBL; AP002561; BAB36800.1; -.
DR   EMBL; M11843; -; NOT_ANNOTATED_CDS.
DR   PIR; A91051; A91051.
DR   PIR; E85895; E85895.
DR   PIR; S23058; S23058.
DR   EcoGene; EG10370; ispG.
DR   HAMAP; MF_00159; -; 1.
DR   InterPro; IPR004588; IspG.
DR   Pfam; PF04551; GcpE; 1.
DR   TIGRFAMs; TIGR00612; ispG_gcpE; 1.
KW   Isoprene biosynthesis; Complete proteome.
SQ   SEQUENCE   372 AA;  40683 MW;  3666FD83D5CD4F9F CRC64;

  Query Match             100.0%;  Score 1866;  DB 1;  Length 372;
  Best Local Similarity   100.0%;  Pred. No. 1.9e-121;
  Matches  372;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MHNQAPIQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MHNQAPIQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVR 60

Qy         61 VSVPTMDAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRM 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VSVPTMDAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRM 120

Qy        121 VVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSV 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 VVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSV 180

Qy        181 KASDVFLAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 KASDVFLAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSL 240

Qy        241 AADPVEEIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDV 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AADPVEEIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDV 300

Qy        301 SIIGCVVNGPGEALVSTLGVTGGNKKSGLYEDGVRKDRLDNNDMIDQLEARIRAKASQLD 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SIIGCVVNGPGEALVSTLGVTGGNKKSGLYEDGVRKDRLDNNDMIDQLEARIRAKASQLD 360

Qy        361 EARRIDVQQVEK 372
              ||||||||||||
Db        361 EARRIDVQQVEK 372


RESULT 2
ISPG_SALTY
ID   ISPG_SALTY     STANDARD;      PRT;   372 AA.
AC   P58671;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase.
GN   ISPG OR GCPE OR STM2523.
OS   Salmonella typhimurium.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Salmonella.
OX   NCBI_TaxID=602;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX   MEDLINE=21534948; PubMed=11677609;
RA   McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA   Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA   Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA   Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA   Waterston R., Wilson R.K.;
RT   "Complete genome sequence of Salmonella enterica serovar Typhimurium
```

```
RT   LT2.";
RL   Nature 413:852-856(2001).
CC   -!- FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate
CC       (ME-2,4cPP) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
CC       (By similarity).
CC   -!- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; sixth step.
CC   -!- SIMILARITY: BELONGS TO THE ISPG FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE008814; AAL21417.1; -.
DR   StyGene; SG?????; ispG.
DR   HAMAP; MF_00159; -; 1.
DR   InterPro; IPR006705; GcpE.
DR   InterPro; IPR004588; IspG.
DR   Pfam; PF04551; GcpE; 1.
DR   TIGRFAMs; TIGR00612; ispG_gcpE; 1.
KW   Isoprene biosynthesis; Complete proteome.
SQ   SEQUENCE   372 AA;  40625 MW;  0486F8D38EC2AC9F CRC64;

  Query Match           98.6%;  Score 1840;  DB 1;  Length 372;
  Best Local Similarity  98.7%;  Pred. No. 1.1e-119;
  Matches  367;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 MHNQAPIQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MHNQAPIQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVR 60

Qy         61 VSVPTMDAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRM 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VSVPTMDAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRM 120

Qy        121 VVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSV 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 VVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSV 180

Qy        181 KASDVFLAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 KASDVFLAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSL 240

Qy        241 AADPVEEIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDV 300
              ||||||||||||||||||:||||||||||||||||||||||||||||||||||||||||||||
Db        241 AADPVEEIKVGFDILKSLRIRARGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDV 300

Qy        301 SIIGCVVNGPGEALVSTLGVTGGNKKSGLYEDGVRKDRLDNNDMIDQLEARIRAKASQLD 360
              ||||||||||||||||||||||||||||||||||||||:||| |||:|||||||||||||
Db        301 SIIGCVVNGPGEALVSTLGVTGGNKKSGLYEDGVRKDRLDNDDMIAQLESRIRAKASQLD 360

Qy        361 EARRIDVQQVEK 372
              ||||||| ||||
Db        361 EARRIDVLQVEK 372


RESULT 3
ISPG_SALTI
ID   ISPG_SALTI     STANDARD;      PRT;   372 AA.
AC   P58670;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase.
GN   ISPG OR GCPE OR STY2768 OR T0333.
OS   Salmonella typhi.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Salmonella.
OX   NCBI_TaxID=601;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CT18;
RX   MEDLINE=21534947; PubMed=11677608;
RA   Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA   Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA   Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
RA   Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA   Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA   Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA   Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA   Whitehead S., Barrell B.G.;
RT   "Complete genome sequence of a multiple drug resistant Salmonella
RT   enterica serovar Typhi CT18.";
RL   Nature 413:848-852(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Ty2 / ATCC 700931;
RX   MEDLINE=22531367; PubMed=12644504;
RA   Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA   Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;
RT   "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RT   and CT18.";
RL   J. Bacteriol. 185:2330-2337(2003).
CC   -!- FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate
CC       (ME-2,4cPP) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
CC       (By similarity).
CC   -!- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; sixth step.
CC   -!- SIMILARITY: BELONGS TO THE ISPG FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL627275; CAD02726.1; -.
DR   EMBL; AE016835; AAO68056.1; -.
DR   HAMAP; MF_00159; -; 1.
DR   InterPro; IPR006705; GcpE.
DR   InterPro; IPR004588; IspG.
DR   Pfam; PF04551; GcpE; 1.
DR   TIGRFAMs; TIGR00612; ispG_gcpE; 1.
KW   Isoprene biosynthesis; Complete proteome.
SQ   SEQUENCE   372 AA;  40626 MW;  3E3D5B13BAAA7916 CRC64;

  Query Match           98.1%;  Score 1831;  DB 1;  Length 372;
  Best Local Similarity  98.1%;  Pred. No. 4.8e-119;
  Matches  365;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 MHNQAPIQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MHNQAPIQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVR 60

Qy         61 VSVPTMDAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRM 120
              ||||||||||||||||||:||||||||||||||||||||||||||||||||||||||||||||
Db         61 VSVPTMDAAEAFKLIKQQVSVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRM 120

Qy        121 VVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSV 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 VVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSV 180

Qy        181 KASDVFLAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 KASDVFLAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSL 240

Qy        241 AADPVEEIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDV 300
              ||||||||||||||||||:||||||||||||||||||||||||||||||||||||||||||||
Db        241 AADPVEEIKVGFDILKSLRIRARGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDV 300

Qy        301 SIIGCVVNGPGEALVSTLGVTGGNKKSGLYEDGVRKDRLDNNDMIDQLEARIRAKASQLD 360
              ||||||||||||||||||||||||||||||||||||||:||| |||:||||| ||||
```

```
Db        301 SIIGCVVNGPGEALVSTLGVTGGNKKSGLYEDGVRKDRLDNDDMIAQLESRIRAKVSQLD 360

Qy        361 EARRIDVQQVEK 372
              ||||||| ||||
Db        361 EARRIDVLQVEK 372


RESULT 4
ISPG_VIBPA
ID   ISPG_VIBPA     STANDARD;      PRT;   372 AA.
AC   Q87S16;
DT   15-SEP-2003 (Rel. 42, Created)
DT   15-SEP-2003 (Rel. 42, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase.
GN   ISPG OR VP0608.
OS   Vibrio parahaemolyticus.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC   Vibrionaceae; Vibrio.
OX   NCBI_TaxID=670;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RIMD 2210633 / Serotype O3:K6;
RX   MEDLINE=22508454; PubMed=12620739;
RA   Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
RA   Iijima Y., Najima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
RA   Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;
RT   "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
RT   distinct from that of V. cholerae.";
RL   Lancet 361:743-749(2003).
CC   -!- FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate
CC       (ME-2,4cPP) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
CC       (By similarity).
CC   -!- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; sixth step.
CC   -!- SIMILARITY: BELONGS TO THE ISPG FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AP005075; BAC58871.1; -.
DR   HAMAP; MF_00159; -; 1.
KW   Isoprene biosynthesis; Complete proteome.
SQ   SEQUENCE   372 AA;  40688 MW;  EAAFAAA1A764DDB8 CRC64;

  Query Match             88.2%;  Score 1645.5;  DB 1;  Length 372;
  Best Local Similarity   87.5%;  Pred. No. 2.8e-106;
  Matches  323;  Conservative   27;  Mismatches   18;  Indels    1;  Gaps    1;

Qy          1 MHNQAPIQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVR 60
              | :::|| |||||||||:||||||||||||||||||||||||| ||:|||:|||||||
Db          1 MQHESPIIRRKSTRIYVGDVPIGDGAPIAVQSMTNTRTTDVEATVAQIRALEKVGADIVR 60

Qy         61 VSVPTMDAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRM 120
              ||||||:|||||||||:|||||||||||||||||||||||||||||||||||||||||
Db         61 VSVPTMEAAEAFKLIKQQVSVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRS 120

Qy        121 VVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSV 180
              |||||||||||||||| |||||||| |||||||:|:|||||||||||||||||||||||
Db        121 VVDCARDKNIPIRIGVNGGSLEKDLQMKYGEPTPEALVESAMRHVDHLDRLNFDQFKVSV 180

Qy        181 KASDVFLAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSL 240
              |||||||||:|||||||:|||||||||||||||||:||||||:|||:|||||||||:||
Db        181 KASDVFLAVDSYRLLAKKIDQPLHLGITEAGGARAGAVKSAVGLGMLLSEGIGDTLRISL 240

Qy        241 AADPVEEIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDV 300
              |||||||||||||||||||:|||||||||||||||||||||||||||||||||||||||
Db        241 AADPVEEIKVGFDILKSLRIRSRGINFIACPSCSRQEFDVIGTVNALEQRLEDIITPMDV 300

Qy        301 SIIGCVVNGPGEALVSTLGVTGGNKKSGLYEDGVR-KDRLDNNDMIDQLEARIRAKASQL 359
              |||||||||||||| || ||: | ||||  |||| | |:| ||||:::||||:|||||| :
Db        301 SIIGCVVNGPGEAEVSHLGLAGSNKKSAFYEDGKRQKERFDNNDLVNQLEAKIRAKASMM 360

Qy        360 DEARRIDVQ 368
              |   ||::: 
Db        361 DSENRIEIK 369


RESULT 5
ISPG_VIBVU
ID   ISPG_VIBVU     STANDARD;      PRT;   372 AA.
AC   Q8DEZ8;
DT   15-SEP-2003 (Rel. 42, Created)
DT   15-SEP-2003 (Rel. 42, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase.
GN   ISPG OR VV10427.
OS   Vibrio vulnificus.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC   Vibrionaceae; Vibrio.
OX   NCBI_TaxID=672;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CMCP6;
RA   Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
RA   Choy H.E.;
RT   "Complete genome sequence of Vibrio vulnificus CMCP6.";
RL   Submitted (DEC-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate
CC       (ME-2,4cPP) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
CC       (By similarity).
CC   -!- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; sixth step.
CC   -!- SIMILARITY: BELONGS TO THE ISPG FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE016798; AAO08950.1; -.
DR   HAMAP; MF_00159; -; 1.
DR   Pfam; PF04551; GcpE; 1.
DR   TIGRFAMs; TIGR00612; ispG_gcpE; 1.
KW   Isoprene biosynthesis; Complete proteome.
SQ   SEQUENCE   372 AA;  40573 MW;  6C73C2B8819B0285 CRC64;

  Query Match             87.4%;  Score 1631.5;  DB 1;  Length 372;
  Best Local Similarity   86.7%;  Pred. No. 2.6e-105;
  Matches  320;  Conservative   29;  Mismatches   19;  Indels    1;  Gaps    1;

Qy          1 MHNQAPIQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVR 60
              |||::|| ||||||||||:||||||||||||||||||||| ||| |||||| |||||||
Db          1 MHNESPIIRRKSTRIYVGDVPIGDGAPIAVQSMTNTRTTDVAATVAQIKALENVGADIVR 60

Qy         61 VSVPTMDAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRM 120
              |||||||||||||||||||::|||||||||||||||||||||||||||||||||||||
Db         61 VSVPTMDAAEAFKLIKQQVSIPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRS 120

Qy        121 VVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSV 180
              |||||||||||||||| |||||||| |||||||:||:|||||||| |||||||||||||
Db        121 VVDCARDKNIPIRIGVNGGSLEKDLQMKYGEPTPEALVESAMRHVDILDRLNFDQFKVSV 180

Qy        181 KASDVFLAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSL 240
              |||||||||:|||||||:|||||||||||||||||:||||||:|||:|||||||||:||
Db        181 KASDVFLAVDSYRLLAKKIDQPLHLGITEAGGARAGAVKSAVGLGMLLAEGIGDTLRISL 240

Qy        241 AADPVEEIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDV 300
              ||:|||||||||||||||||||||||||:||||||| |||||:||||||||||||||||
Db        241 AANPVEEIKVGFDILKSLRIRSRGINFIACPSCSRQEFDVISTVNALEERLEDIITPMDV 300
```

```
Qy        301 SIIGCVVNGPGEALVSTLGVTGGNKKSGLYEDGVR-KDRLDNNDMIDQLEARIRAKASQL 359
              ||||||||||||| || ||: | |||| |||| | |:| || |:::|||||:||||||::
Db        301 SIIGCVVNGPGEAEVSHLGLAGSNKKSAFYEDGKRQKERFDNEDLVNQLEAKIRAKAARM 360

Qy        360 DEARRIDVQ 368
              ||: |||::
Db        361 DESNRIDIK 369


RESULT 6
ISPG_PROST
ID   ISPG_PROST     STANDARD;      PRT;   365 AA.
AC   P72241;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase.
GN   ISPG OR AARC.
OS   Providencia stuartii.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Providencia.
OX   NCBI_TaxID=588;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PR50;
RX   MEDLINE=97234638; PubMed=9079912;
RA   Rather P.N., Solinsky K.A., Paradise M.R., Parojcic M.M.;
RT   "aarC, an essential gene involved in density-dependent regulation of
RT   the 2'-N-acetyltransferase in Providencia stuartii.";
RL   J. Bacteriol. 179:2267-2273(1997).
CC   -!- FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate
CC       (ME-2,4cPP) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
CC       (By similarity). Involved in density-dependent regulation of 2'-N-
CC       acetyltransferase.
CC   -!- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; sixth step.
CC   -!- SIMILARITY: BELONGS TO THE ISPG FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U67933; AAB51469.1; -.
DR   HAMAP; MF_00159; -; 1.
DR   InterPro; IPR004588; IspG.
DR   Pfam; PF04551; GcpE; 1.
DR   TIGRFAMs; TIGR00612; ispG_gcpE; 1.
KW   Isoprene biosynthesis.
SQ   SEQUENCE   365 AA;  39840 MW;  4523147980480ECC CRC64;

  Query Match             87.0%;  Score 1622.5;  DB 1;  Length 365;
  Best Local Similarity   88.4%;  Pred. No. 1e-104;
  Matches  320;  Conservative   24;  Mismatches   17;  Indels    1;  Gaps    1;

Qy          1 MHNQAPIQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVR 60
              |||::||:|||||||||||||||||||||||||||||||||||||||| ||::|||||| ||||
Db          1 MHNESPIKRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVRQIQSLERVGVDIVR 60

Qy         61 VSVPTMDAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRM 120
              |||||||||||||||||:|||||||||||||||:|||||||||||||||||||:|||||
Db         61 VSVPTMDAAEAFKLIKQRVNVPLVADIHFDYRIAMKVAEYGVDCLRINPGNIGSEERIRQ 120

Qy        121 VVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSV 180
              ||| ||  ||||||||| ||||||:||||||||:||:|||||||| ||||||||||||
Db        121 VVDSARHHNIPIRIGVNGGSLEKDIQEKYGEPTPEALVESAMRHVDILDRLNFDQFKVSV 180

Qy        181 KASDVFLAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSL 240
              ||||||||| |||||::|||||||||||||||:||||||||:||:|||||||:||
Db        181 KASDVFLAVGSYRLLAQKIDQPLHLGITEAGGARSGSVKSAIGLGMLLAEGIGDTLRISL 240

Qy        241 AADPVEEIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDV 300
              |||||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AADPVEEVKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDV 300

Qy        301 SIIGCVVNGPGEALVSTLGVTGGNKKSGLYEDGVR-KDRLDNNDMIDQLEARIRAKASQL 359
              |||||||||||||| |||||| |   ||| |||||| |:| ||:::|||||||:|||||: |
Db        301 SIIGCVVNGPGEAEVSTLGVAGAKTKSGFYEDGVRKKERFDNDNIIDQLEAKIRAKAAML 360

Qy        360 DE 361
              ||
Db        361 DE 362


RESULT 7
ISPG_YERPE
ID   ISPG_YERPE     STANDARD;      PRT;   375 AA.
AC   P58672;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase.
GN   ISPG OR GCPE OR YPO2879 OR Y1353.
OS   Yersinia pestis.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Yersinia.
OX   NCBI_TaxID=632;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CO-92 / Biovar Orientalis;
RX   MEDLINE=21470413; PubMed=11586360;
RA   Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA   Prentice M.B., Sebaihia M., James K.D., Churcher C., Mungall K.L.,
RA   Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
RA   Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA   Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
RA   Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA   Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT   "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL   Nature 413:523-527(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KIM5 / Biovar Mediaevalis;
RX   MEDLINE=22137863; PubMed=12142430;
RA   Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
RA   Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
RA   Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
RA   Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,
RA   Perry R.D.;
RT   "Genome sequence of Yersinia pestis KIM.";
RL   J. Bacteriol. 184:4601-4611(2002).
CC   -!- FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate
CC       (ME-2,4cPP) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
CC       (By similarity).
CC   -!- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; sixth step.
CC   -!- SIMILARITY: BELONGS TO THE ISPG FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ414154; CAC92130.1; -.
DR   EMBL; AE013738; AAM84926.1; -.
DR   PIR; AG0350; AG0350.
DR   HAMAP; MF_00159; -; 1.
DR   InterPro; IPR006705; GcpE.
DR   InterPro; IPR004588; IspG.
DR   Pfam; PF04551; GcpE; 1.
DR   TIGRFAMs; TIGR00612; ispG_gcpE; 1.
```

```
KW   Isoprene biosynthesis; Complete proteome.
SQ   SEQUENCE   375 AA;  40797 MW;  860971872B1C0536 CRC64;

  Query Match             87.0%;  Score 1622.5;  DB 1;  Length 375;
  Best Local Similarity   86.8%;  Pred. No. 1.1e-104;
  Matches  323;  Conservative   19;  Mismatches   29;  Indels    1;  Gaps    1;

Qy          1 MHNQAPIQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVR 60
              ||| :|| |||||||||||| ||||||||||||||||||:||||:||| |||||||| ||||
Db          1 MHNGSPIIRRKSTRIYVGKVPIGDGAPIAVQSMTNTKTTDVDATVAQIKALERVGVDIVR 60

Qy         61 VSVPTMDAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRM 120
              ||:|||||||||||||||  ||||||||||||||||||||||||||||||||||||| |||
Db         61 VSIPTMDAAEAFKLIKQQSTVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNESRIRE 120

Qy        121 VVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSV 180
              || |||| |||||||:| |||||||:||||||||:|||||||||| ||||||||||||||
Db        121 VVACARDYNIPIRIGINGGSLEKDIQEKYGEPTPEALLESAMRHVDILDRLNFDQFKVSV 180

Qy        181 KASDVFLAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSL 240
              |||||||||| |||||||||: ||||||||||||||:|||||||||||||||||||:||
Db        181 KASDVFLAVNSYRLLAKQINNPLHLGITEAGGARSGSVKSAIGLGLLLSEGIGDTLRISL 240

Qy        241 AADPVEEIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDV 300
              |||||||:|||||||||||||:|||||||||||||||||||||||||||:||||||
Db        241 AADPVEEVKVGFDILKSLRIRARGINFIACPTCSRQEFDVIGTVNALEQRLEDLITPMDV 300

Qy        301 SIIGCVVNGPGEALVSTLGVTGGNKKSGLYEDGVR-KDRLDNNDMIDQLEARIRAKASQL 359
              |||||||||||||||||8|||||    || |||||| |:| ||  |||||||:|||||| |
Db        301 SIIGCVVNGPGEALVSTLGVTGARNHSGFYEDGVRQKERFDNKAMIDQLEAKIRAKASIL 360

Qy        360 DEARRIDVQQVE 371
              |   || : |::
Db        361 DANNRIVINQLD 372


RESULT 8
ISPG_VIBCH
ID   ISPG_VIBCH     STANDARD;      PRT;   376 AA.
AC   Q9KTX1;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase.
GN   ISPG OR VC0759.
OS   Vibrio cholerae.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC   Vibrionaceae; Vibrio.
OX   NCBI_TaxID=666;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=El Tor N16961 / Serotype O1;
RX   MEDLINE=20406833; PubMed=10952301;
RA   Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA   Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA   Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA   Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA   McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA   Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA   Fraser C.M.;
RT   "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT   cholerae.";
RL   Nature 406:477-483(2000).
CC   -!- FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate
CC       (ME-2,4cPP) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
CC       (By similarity).
CC   -!- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; sixth step.
CC   -!- SIMILARITY: BELONGS TO THE ISPG FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE004161; AAF93924.1; -.
DR   PIR; F82283; F82283.
DR   TIGR; VC0759; -.
DR   HAMAP; MF_00159; -; 1.
DR   InterPro; IPR004588; IspG.
DR   Pfam; PF04551; GcpE; 1.
DR   TIGRFAMs; TIGR00612; ispG_gcpE; 1.
KW   Isoprene biosynthesis; Complete proteome.
SQ   SEQUENCE   376 AA;  40862 MW;  34E4144CBB7ADAAA CRC64;

  Query Match             86.9%;  Score 1621.5;  DB 1;  Length 376;
  Best Local Similarity   85.3%;  Pred. No. 1.3e-104;
  Matches  318;  Conservative   33;  Mismatches   21;  Indels    1;  Gaps    1;

Qy          1 MHNQAPIQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVR 60
              | :::||:|| |||||||:||||||||||||||||||| ||| |||:||:|||||||
Db          3 MQHESPIKRRPSTRIYVGDVPIGDGAPIAVQSMTNTRTTDVAATVAQIKSLEKVGADIVR 62

Qy         61 VSVPTMDAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRM 120
              ||||||:|||||||||||:|||||||||||||||||||||||||||||||||| |||
Db         63 VSVPTMEAAEAFKLIKQQVSVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEARIRS 122

Qy        121 VVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSV 180
              |||||||| |||||||| |||||||:||:|||||||| ||||||||||||||||||||
Db        123 VVDCARDKGIPIRIGVNGGSLEKDLQLKYGEPTPEALVESAMRHVDILDRLNFDQFKVSV 182

Qy        181 KASDVFLAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSL 240
              |||||||||:|||||||:|||||||||||||:|:||||:|||:||:|||||||:||
Db        183 KASDVFLAVDSYRLLAKKIDQPLHLGITEAGGARAGSVKSAVGLGMLLAEGIGDTLRISL 242

Qy        241 AADPVEEIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDV 300
              |||||||||||||||||||||||||||:|||||||||||||||||||::|||||
Db        243 AADPVEEIKVGFDILKSLRIRSRGINFIACPSCSRQEFDVIGTVNALEQRLEDVLTPMDV 302

Qy        301 SIIGCVVNGPGEALVSTLGVTGGNKKSGLYEDGVR-KDRLDNNDMIDQLEARIRAKASQL 359
              |||||||||||| || ||: | ||||  |||||| |:| ||:|:: ||||:|||||::|
Db        303 SIIGCVVNGPGEAEVSHLGLAGSNKKSAFYEDGVRQKERFDNDDLVAQLEAKIRAKAARL 362

Qy        360 DEARRIDVQQVEK 372
              ||  |||:: ||:
Db        363 DEKNRIDIKHVEQ 375


RESULT 9
ISPG_HAEIN
ID   ISPG_HAEIN     STANDARD;      PRT;   368 AA.
AC   P44667;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase.
GN   ISPG OR GCPE OR HI0368.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC   Pasteurellaceae; Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Rd / KW20 / ATCC 51907;
RX   MEDLINE=95350630; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
```

```
RT   "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT   Rd.";
RL   Science 269:496-512(1995).
CC   -!- FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate
CC       (ME-2,4cPP) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
CC       (By similarity).
CC   -!- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; sixth step.
CC   -!- SIMILARITY: BELONGS TO THE ISPG FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U32721; AAC22026.1; -.
DR   PIR; H64063; H64063.
DR   TIGR; HI0368; -.
DR   HAMAP; MF_00159; -; 1.
DR   InterPro; IPR004588; IspG.
DR   Pfam; PF04551; GcpE; 1.
DR   TIGRFAMs; TIGR00612; ispG_gcpE; 1.
KW   Isoprene biosynthesis; Complete proteome.
SQ   SEQUENCE   368 AA;  40116 MW;  DC99EA15A50ED01B CRC64;

  Query Match             85.9%;  Score 1603.5;  DB 1;  Length 368;
  Best Local Similarity   88.2%;  Pred. No. 2.1e-103;
  Matches  320;  Conservative   20;  Mismatches   22;  Indels    1;  Gaps    1;

Qy          4 QAPIQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVRVSV 63
              |  |:||:||:|||||||||  ||||||||||||||||||| |||:||||||||||||||
Db          5 QPTIKRRESTKIYVGNVPIGGDAPIAVQSMTNTRTTDVEATVAQIKSLERVGADIVRVSV 64

Qy         64 PTMDAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRMVVD 123
              |||||||||| ||||||||||||||||||||||||||||||||||||| |:|:| |||
Db         65 PTMDAAEAFKQIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGREDRVRAVVD 124

Qy        124 CARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSVKAS 183
              ||||||||||||||||||||||||||||:||||||:|||: ||||||||||||||||
Db        125 CARDKNIPIRIGVNAGSLEKDLQEKYGEPTPEALLESALRHVEILDRLNFDQFKVSVKAS 184

Qy        184 DVFLAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSLAAD 243
              |||||||||||||  |||||||||||||:||||||:|||:||:|||||||||||||||
Db        185 DVFLAVESYRLLAKAIKQPLHLGITEAGGARAGAVKSAVGLGMLLAEGIGDTLRVSLAAD 244

Qy        244 PVEEIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDVSII 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        245 PVEEIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDVSII 304

Qy        304 GCVVNGPGEALVSTLGVTGGNKKSGLYEDGVR-KDRLDNNDMIDQLEARIRAKASQLDEA 362
              |||||||||||| |||||||||| | || | |:| || |:::||||:|||| :: |
Db        305 GCVVNGPGEALVSDLGVTGGNKKSGYYLDGERQKERFDNEDIVNQLEAKIRAKVARQDPK 364

Qy        363 RRI 365
              ||
Db        365 NRI 367


RESULT 10
ISPG_PASMU
ID   ISPG_PASMU     STANDARD;      PRT;   367 AA.
AC   P57987;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase.
GN   ISPG OR GCPE OR PM2010.
OS   Pasteurella multocida.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC   Pasteurellaceae; Pasteurella.
OX   NCBI_TaxID=747;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Pm70;
RX   MEDLINE=21145866; PubMed=11248100;
RA   May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT   "Complete genomic sequence of Pasteurella multocida Pm70.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC   -!- FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate
CC       (ME-2,4cPP) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
CC       (By similarity).
CC   -!- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; sixth step.
CC   -!- SIMILARITY: BELONGS TO THE ISPG FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE006237; AAK04094.1; -.
DR   HAMAP; MF_00159; -; 1.
DR   InterPro; IPR004588; IspG.
DR   Pfam; PF04551; GcpE; 1.
DR   TIGRFAMs; TIGR00612; ispG_gcpE; 1.
KW   Isoprene biosynthesis; Complete proteome.
SQ   SEQUENCE   367 AA;  39969 MW;  1918C5702642AEE8 CRC64;

  Query Match             85.7%;  Score 1599.5;  DB 1;  Length 367;
  Best Local Similarity   88.9%;  Pred. No. 4e-103;
  Matches  320;  Conservative   18;  Mismatches   21;  Indels    1;  Gaps    1;

Qy          7 IQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVRVSVPTM 66
              |:||:||:|||||||||  |||||||||||||||||| ||||||||||||||||||||
Db          7 IKRRESTKIYVGNVPIGGDAPIAVQSMTNTRTTDVEATVAQIKALERVGADIVRVSVPTM 66

Qy         67 DAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRMVVDCAR 126
              |||||||||||||||||||||||||||||||||||||||||||||||||| |:||| ||||||
Db         67 DAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGREDRIRAVVDCAR 126

Qy        127 DKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSVKASDVF 186
              | ||||||||||||||||||||||||||:||||||:|||: ||||||||||||||||||
Db        127 DNNIPIRIGVNAGSLEKDLQEKYGEPTPEALLESALRHVEILDRLNFDQFKVSVKASDVF 186

Qy        187 LAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSLAADPVE 246
              |||||||||| | |||||||||||| |:|||||:|||:||:|||||||:|||||||||
Db        187 LAVESYRLLAKAIKQPLHLGITEAGGFRAGAVKSAVGLGMLLAEGIGDTLRISLAADPVE 246

Qy        247 EIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDVSIIGCV 306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        247 EIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDVSIIGCV 306

Qy        307 VNGPGEALVSTLGVTGGNKKSGLYEDGVR-KDRLDNNDMIDQLEARIRAKASQLDEARRI 365
              |||||||||| |||||||||| | || | |:| ||:|:|:||||:|||| :  |   ||
Db        307 VNGPGEALVSDLGVTGGNKKSGYYLDGERQKERFDNDDLINQLEAKIRAKVAAQDPKNRI 366


RESULT 11
ISPG_PSEAE
ID   ISPG_PSEAE     STANDARD;      PRT;   371 AA.
AC   Q9HXJ4;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase.
GN   ISPG OR PA3803.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC   Pseudomonadaceae; Pseudomonas.
OX   NCBI_TaxID=287;
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RC   STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE=20437337; PubMed=10984043;
RA   Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA   Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA   Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA   Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA   Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA   Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT   "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT   opportunistic pathogen.";
RL   Nature 406:959-964(2000).
CC   -!- FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate
CC       (ME-2,4cPP) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
CC       (By similarity).
CC   -!- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; sixth step.
CC   -!- SIMILARITY: BELONGS TO THE ISPG FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE004798; AAG07190.1; -.
DR   PIR; F83171; F83171.
DR   HAMAP; MF_00159; -; 1.
DR   InterPro; IPR004588; IspG.
DR   Pfam; PF04551; GcpE; 1.
DR   TIGRFAMs; TIGR00612; ispG_gcpE; 1.
KW   Isoprene biosynthesis; Complete proteome.
SQ   SEQUENCE   371 AA;  40056 MW;  F33FC1FCEF017117 CRC64;

  Query Match             73.6%;  Score 1372.5;  DB 1;  Length 371;
  Best Local Similarity   73.8%;  Pred. No. 1.7e-87;
  Matches  267;  Conservative   46;  Mismatches   48;  Indels    1;  Gaps    1;

Qy          1 MHNQAPIQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVR 60
              :|: :|| |||| :|:|||||:|  |||||||||| | || ||| ||: ||  ||||||
Db          3 IHSASPIIRRKSRKIWVGNVPVGGDAPIAVQSMTNTETCDVAATVAQIRRLEDAGADIVR 62

Qy         61 VSVPTMDAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRM 120
              |||| |||||||  |||||||||||||||||||||:||| |||||||||||| |:|::
Db         63 VSVPDMDAAEAFGKIKQQVNVPLVADIHFDYRIALRVAELGVDCLRINPGNIGREDRVKA 122

Qy        121 VVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSV 180
              ||| ||::|||||||||||||||||:|||||||:|||||||||||||||:|:|  |||||
Db        123 VVDAARERNIPIRIGVNAGSLEKDLQKKYGEPTPEALLESAMRHVDHLDKLDFQNFKVSV 182

Qy        181 KASDVFLAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSL 240
              ||||||:|| :|||||8||:||||||||||| ||| ||||:|||:||:|||||||:|:||
Db        183 KASDVFMAVAAYRLLARQIEQPLHLGITEAGGLRSGTVKSAVGLGMLLAEGIGDTIRISL 242

Qy        241 AADPVEEIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDV 300
              |||||||||||||||||| :|||||||||||:|||| ||| |:| || ||||:: ||||
Db        243 AADPVEEIKVGFDILKSLHLRSRGINFIACPSCSRQNFDVVKTMNELEGRLEDLLVPMDV 302

Qy        301 SIIGCVVNGPGEALVSTLGVTGGNKKSGLYEDGVRKDRLDNNDMIDQLEARIRAKASQLD 360
              ::|||||||||||  :|:|:|||     :| ||   :| |::::|:|| || ||::
Db        303 AVIGCVVNGPGEAKEAHVGLTGGTPNL-VYIDGKPSQKLTNDNLVDELERLIRQKAAEKA 361

Qy        361 EA 362
              ||
Db        362 EA 363


RESULT 12
ISPG_BUCAP
ID   ISPG_BUCAP     STANDARD;      PRT;   367 AA.
AC   Q8K9P4;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase.
GN   ISPG OR GCPE OR BUSG276.
OS   Buchnera aphidicola (subsp. Schizaphis graminum).
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Buchnera.
OX   NCBI_TaxID=98794;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22084549; PubMed=12089438;
RA   Tamas I., Klasson L., Canbaeck B., Naeslund A.K., Eriksson A.-S.,
RA   Wernegreen J.J., Sandstroem J.P., Moran N.A., Andersson S.G.E.;
RT   "50 million years of genomic stasis in endosymbiotic bacteria.";
RL   Science 296:2376-2379(2002).
CC   -!- FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate
CC       (ME-2,4cPP) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
CC       (By similarity).
CC   -!- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; sixth step.
CC   -!- SIMILARITY: BELONGS TO THE ISPG FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE014104; AAM67834.1; -.
DR   HAMAP; MF_00159; -; 1.
DR   InterPro; IPR006705; GcpE.
DR   InterPro; IPR004588; IspG.
DR   Pfam; PF04551; GcpE; 1.
DR   TIGRFAMs; TIGR00612; ispG_gcpE; 1.
KW   Isoprene biosynthesis; Complete proteome.
SQ   SEQUENCE   367 AA;  40440 MW;  4CF6A0FAF017661C CRC64;

  Query Match             68.9%;  Score 1285.5;  DB 1;  Length 367;
  Best Local Similarity   66.7%;  Pred. No. 1.6e-81;
  Matches  244;  Conservative   64;  Mismatches   57;  Indels    1;  Gaps    1;

Qy          1 MHNQAPIQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVR 60
              |:    |:|||| |||||||||||||||:|||||||:||::| |: ||  |::|| ||||
Db          1 MNKYKNIKRRKSNRIYVGNVPIGDGAPISVQSMTNTQTTNIEETIKQIIKLKKVGVDIVR 60

Qy         61 VSVPTMDAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRM 120
              :||||::|||:|||||  |:|||:|||||||||:|:|  :|| ||||||||||: ||
Db         61 ISVPTLEAAESFKIIKLNVDVPLIADIHFDYRLAIKSIKYGADCLRINPGNIGNKRRILD 120

Qy        121 VVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSV 180
              :|:||:|||||||||||||||| || :||  | |:||:|||:||:::|| |||:||||||
Db        121 IVNCAKDKNIPIRIGVNAGSLENDLLKKYKSPIPEALVESAIRHIEYLDSLNFNQFKVSV 180

Qy        181 KASDVFLAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSL 240
              | |||| |:|:  :|||:  ||:|:||||:|  |:| |||:||:  ||  |||||||:||
Db        181 KTSDVFSAIEANEILAKKTVQPIHIGITESGALRNGIVKSSIGITSLLLSGIGDTLRISL 240

Qy        241 AADPVEEIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDV 300
              |||||||:|||:||||:| || ||:|||||||||||||:||  || ||: |||: |||:|
Db        241 AADPVEEVKVGYDILKTLGIRFRGVNFIACPTCSRQEFNVIDVVNQLEKNLEDLSTPMNV 300

Qy        301 SIIGCVVNGPGEALVSTLGVTGGNKKSGLYEDGVR-KDRLDNNDMIDQLEARIRAKASQL 359
              |||||:||| ||| ||||||: ||:| | ||:||:| |::| | ::| :|| :|| ||  |
Db        301 SIIGCIVNGIGEAKVSTLGIVGGSKTSALYKDGIRQKNKLKNQEIIKELEIKIRKKAKSL 360

Qy        360 DEARRI 365
              |: ::|
Db        361 DKLKKI 366


RESULT 13
ISPG_BUCAI
ID   ISPG_BUCAI     STANDARD;      PRT;   368 AA.
```

```
AC   P57374;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase.
GN   ISPG OR GCPE OR BU287.
OS   Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OS   symbiotic bacterium).
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Buchnera.
OX   NCBI_TaxID=118099;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Tokyo 1998;
RX   MEDLINE=20445173; PubMed=10993077;
RA   Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT   "Genome sequence of the endocellular bacterial symbiont of aphids
RT   Buchnera sp. APS.";
RL   Nature 407:81-86(2000).
CC   -!- FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate
CC       (ME-2,4cPP) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
CC       (By similarity).
CC   -!- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; sixth step.
CC   -!- SIMILARITY: BELONGS TO THE ISPG FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AP001118; BAB12997.1; -.
DR   HSSP; P49058; 1EEP.
DR   HAMAP; MF_00159; -; 1.
DR   InterPro; IPR004588; IspG.
DR   Pfam; PF04551; GcpE; 1.
DR   TIGRFAMs; TIGR00612; ispG_gcpE; 1.
KW   Isoprene biosynthesis; Complete proteome.
SQ   SEQUENCE   368 AA;  40869 MW;  8BC32F495954C1DE CRC64;

  Query Match           68.7%;  Score 1282;  DB 1;  Length 368;
  Best Local Similarity  66.1%;  Pred. No. 2.9e-81;
  Matches  242;  Conservative   65;  Mismatches   57;  Indels    2;  Gaps    2;

Qy          1 MHNQAP-IQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIV 59
              |:|:   | |||| ||||| | ||: |||:||||||||||::  |:|||  |::|| |||
Db          1 MNNKCKIINRRKSDRIYVGKVAIGNNAPISVQSMTNTRTTNISETINQILELQKVGVDIV 60

Qy         60 RVSVPTMDAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIR 119
              |:|:| : |||:|| ||:| ||||:||||||||:||:  :|| ||||||||||||: |:
Db         61 RISIPNLKAAESFKEIKKQTNVPLIADIHFDYRLALQAIKYGADCLRINPGNIGNKRRVS 120

Qy        120 MVVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVS 179
               ::  |:|:|||||||||||||||||: :||  ||| ||:|||||::: | |||:|||||
Db        121 EIISYAKDENIPIRIGVNAGSLEKDILKKYKIPTPDALVESAMRHIEYFDALNFNQFKVS 180

Qy        180 VKASDVFLAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVS 239
              |||||||||:||||:| |:| ||||:||||:|| |:| |||:||: ||| |||||||:|||
Db        181 VKASDVFLAIESYRMLGKKITQPLHIGITESGGLRNGTVKSSIGIALLLLEGIGDTIRVS 240

Qy        240 LAADPVEEIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMD 299
              ||| | ||:|||:|||| | :|:|||||||||||||||||| ||| ||: |||| ||:|
Db        241 LAAHPTEEVKVGYDILKVLSLRARGINFIACPTCSRQEFDVINTVNQLEKNLEDISTPID 300

Qy        300 VSIIGCVVNGPGEALVSTLGVTGGNKKSGLYEDGVR-KDRLDNNDMIDQLEARIRAKASQ 358
              |||||||||| ||: ::||||: | :||| |||||| |::: | ::|:::| :|| |  :
Db        301 VSIIGCVVNGIGESKIATLGLAGSHKKSAFYEDGVRQKEKIKNEEIIEKMEIKIRKKIDK 360

Qy        359 LDEARR 364
              |: :::
Db        361 LNNSKK 366
```

```
RESULT 14
ISPG_WIGBR
ID   ISPG_WIGBR     STANDARD;      PRT;   366 AA.
AC   Q8D1Y3;
DT   15-SEP-2003 (Rel. 42, Created)
DT   15-SEP-2003 (Rel. 42, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase.
GN   ISPG OR GCPE OR WIGBR5730.
OS   Wigglesworthia glossinidia brevipalpis.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Wigglesworthia.
OX   NCBI_TaxID=36870;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22297718; PubMed=12219091;
RA   Akman L., Yamashita A., Watanabe H., Oshima K., Shiba T., Hattori M.,
RA   Aksoy S.;
RT   "Genome sequence of the endocellular obligate symbiont of tsetse
RT   flies, Wigglesworthia glossinidia.";
RL   Nat. Genet. 32:402-407(2002).
CC   -!- FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate
CC       (ME-2,4cPP) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
CC       (By similarity).
CC   -!- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; sixth step.
CC   -!- SIMILARITY: BELONGS TO THE ISPG FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB063522; BAC24719.1; -.
DR   HAMAP; MF_00159; -; 1.
DR   InterPro; IPR006705; GcpE.
DR   InterPro; IPR004588; IspG.
DR   Pfam; PF04551; GcpE; 1.
DR   TIGRFAMs; TIGR00612; ispG_gcpE; 1.
KW   Isoprene biosynthesis; Complete proteome.
SQ   SEQUENCE   366 AA;  40407 MW;  159E0169E8FC8DED CRC64;

  Query Match           65.7%;  Score 1226.5;  DB 1;  Length 366;
  Best Local Similarity  65.9%;  Pred. No. 1.9e-77;
  Matches  234;  Conservative   62;  Mismatches   58;  Indels    1;  Gaps    1;

Qy          7 IQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVRVSVPTM 66
              | |||| :||: ||||| :||:||||||| |||: :|::|| |:: |||||||:||:
Db          7 IIRRKSKKIYIKNVPIGGDSPISVQSMTNTCTTDINSTISQINKLQKAGADIVRVSIPTL 66

Qy         67 DAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRMVVDCAR 126
              :|||:||:||: |::|:|||||||||||| |||| ||||||||||   || || |:
Db         67 EAAESFKIIKRNVSIPIVADIHFDYRIALKAAEYGADCLRINPGNIGKLNRIISVVSTAK 126

Qy        127 DKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSVKASDVF 186
              :| :|||||||:||||||:: |||   |::| |||||||: |::|||| ||||||:|||
Db        127 EKKLPIRIGVNSGSLEKDIENKYGINNPKSLFESAMRHVNILEKLNFDMFKVSVKSSDVL 186

Qy        187 LAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSLAADPVE 246
              |:||:||| :||||||||||:|    |::|:||:|||||||||||||||||||:|
Db        187 TCVQSYKLLASKIDQPLHLGITESGSMLHGSIKSSIGIGLLLSEGIGDTLRVSLAADPIE 246

Qy        247 EIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDVSIIGCV 306
              |:|||| ||:|| || |||||||||||||||||  || ||:||||:||||:|||:
Db        247 EVKVGFSILRSLNIRKRGINFIACPTCSRQEFDVINVVNVLEKRLEDVITPMNVSVIGCM 306

Qy        307 VNGPGEALVSTLGVTGGNKKSGLYEDGVR-KDRLDNNDMIDQLEARIRAKASQLD 360
              ||| |||  : :|::|   || |:|:|:|  ::::| ::|||:||  || |   |:
Db        307 VNGLGEANRADIGISGSRNKSILFENGLRNNNKINNEEIIDKLEKYIRKKVKILN 361
```

```
RESULT 15
ISPG_CAUCR
ID   ISPG_CAUCR     STANDARD;      PRT;   383 AA.
AC   Q9A9W0;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase.
GN   ISPG OR CC0851.
OS   Caulobacter crescentus.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacterales;
OC   Caulobacteraceae; Caulobacter.
OX   NCBI_TaxID=155892;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 19089 / CB15;
RX   MEDLINE=21173698; PubMed=11259647;
RA   Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA   Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA   Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA   DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA   Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA   Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA   Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT   "Complete genome sequence of Caulobacter crescentus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
CC   -!- FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate
CC       (ME-2,4cPP) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
CC       (By similarity).
CC   -!- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; sixth step.
CC   -!- SIMILARITY: BELONGS TO THE ISPG FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AE005761; AAK22836.1; -.
DR   PIR; H87354; H87354.
DR   TIGR; CC0851; -.
DR   HAMAP; MF_00159; -; 1.
DR   InterPro; IPR004588; IspG.
DR   Pfam; PF04551; GcpE; 1.
DR   TIGRFAMs; TIGR00612; ispG_gcpE; 1.
KW   Isoprene biosynthesis; Complete proteome.
SQ   SEQUENCE   383 AA;  40788 MW;  8468EB400E338194 CRC64;

  Query Match             55.9%;  Score 1042.5;  DB 1;  Length 383;
  Best Local Similarity   57.6%;  Pred. No. 9.2e-65;
  Matches  208;  Conservative   57;  Mismatches   95;  Indels    1;  Gaps    1;

Qy          7 IQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVRVSVPTM 66
              | ||:| :| ||:| :||  |||:||||||| |:|   ||: ||: ||  |||||||| | :
Db         14 ITRRQSRKIRVGSVEVGGDAPISVQSMTNTLTSDAAATLEQIRQLEEAGADIVRVSCPDV 73

Qy         67 DAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRMVVDCAR 126
              :: |||  | ::   ||||||| |: ::  |: |  ||||||||||: :|:| |:  ||
Db         74 ESTAAFKTIAREAKVPLVADIHFHYKRGIEAAQAGAACLRINPGNIGSPDRVRDVIQAAR 133

Qy        127 DKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSVKASDVF 186
              |    :||||||||||:: ||||||  | ::|||: |     |      :| :||:|||||| |
Db        134 DHGCSMRIGVNAGSLERELLEKYGEPCPDAMVESALNHARILQDHDFHEFKISVKASDPF 193

Qy        187 LAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSLAADPVE 246
              : | :|  ||: || ||||:||||   |:| ||||||:| :|   |||||:||||||||||||
Db        194 MTVAAYYQLAEAIDCPLHLGVTEAGATRTGTVKSAIGIGAMLWAGIGDTIRVSLAADPVE 253

Qy        247 EIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDVSIIGCV 306
              ||||||||||| :| ||:| ||||:|:|| |:|| || |||:|| | ||| :||||||
Db        254 EIKVGFDILKSLGLRHRGVNIIACPSCARQGFNVIKTVEALEERLAHISTPMSLSIIGCV 313

Qy        307 VNGPGEALVSTLGVTGGNKKSGL-YEDGVRKDRLDNNDMIDQLEARIRAKASQLDEARRI 365
              ||||||||:: :| |||   :|: |  |     :  |  |||  :    :  ||:::  |:
Db        314 VNGPGEALMTDIGFTGGGAGAGMVYMAGKPDHKQSNEGMIDHIVDLVEKKAAEIQAAKAQ 373

Qy        366 D 366
              |
Db        374 D 374

Search completed: August 12, 2003, 09:54:29
Job time : 25 secs
```

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 14, 2003, 21:54:20 ; Search time 2622 Seconds
(without alignments)
3448.232 Million cell updates/sec

Title: US-09-921-992-78
Perfect score: 1866
Sequence: 1 MHNQAPIQRRKSTRIYVGNV.....RAKASQLDEARRIDVQQVEK 372

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09921992/runat_12082003_094912_17379/app_query.fasta_1.519
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09921992_@CGN_1_1_2810_@runat_12082003_094912_17379 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	820.5	44.0	1323	29	BZ575759	BZ575759 msh2_4614
2	379	20.3	828	14	CB669733	CB669733 OSJNEe02H
3	348.5	18.7	853	14	CB619369	CB619369 OSIIEa03O
4	326.5	17.5	1442	11	AY104363	AY104363 Zea mays
c 5	321.5	17.2	1268	29	BZ569685	BZ569685 pacs2-164
6	315.5	16.9	579	12	BI920888	BI920888 EST540823
7	313	16.8	720	28	BH235010	BH235010 PH_05.x S
c 8	307.5	16.5	658	28	AZ570993	AZ570993 281PvD07
9	276.5	14.8	852	14	CB628479	CB628479 OSIIEb04B
10	271	14.5	602	14	CA022320	CA022320 HZ42M11r
11	269.5	14.4	610	12	BJ481080	BJ481080 BJ481080
12	265	14.2	706	10	BG591263	BG591263 EST499105
13	264	14.1	489	6	AU186794	Au186794 Porphyra
14	262	14.0	601	9	AW042702	AW042702 ST23E04 P
15	261	14.0	603	12	BJ551332	BJ551332 BJ551332
16	260	13.9	558	9	AV917069	AV917069 AV917069
17	260	13.9	608	12	BJ465443	BJ465443 BJ465443
18	260	13.9	616	9	AV934187	AV934187 AV934187
19	259	13.9	484	13	BQ080994	BQ080994 san12g09.
20	259	13.9	651	9	AU252386	AU252386 AU252386
21	257	13.8	640	13	BU091269	BU091269 st70g12.y
22	252	13.5	579	28	AZ523878	AZ523878 224PbE02
c 23	250.5	13.4	351	13	BU004160	BU004160 QGG37N12.
24	244	13.1	465	9	AI482869	AI482869 EST242192
25	236	12.6	576	13	BQ975044	BQ975044 QHI18D05.
26	232.5	12.5	545	9	AV626844	AV626844 AV626844
27	229	12.3	532	10	BE924238	BE924238 EST428007
28	229	12.3	628	13	BU048059	BU048059 PP_LEa003
29	229	12.3	680	12	BG890681	BG890681 EST516532
30	229	12.3	700	12	BI933447	BI933447 EST553336
31	229	12.3	728	10	BG525868	BG525868 53-47-1-2
32	228	12.2	423	13	BQ862800	BQ862800 QGC22B19.
33	225	12.1	792	14	CB620208	CB620208 OSIIEa05D
34	225	12.1	821	14	CB291771	CB291771 UCRCS01_0
35	225	12.1	821	14	CB628414	CB628414 OSIIEb03P
36	225	12.1	1127	29	BZ577824	BZ577824 msh2_5589
37	224	12.0	687	14	CA781720	CA781720 046F07AF
38	223.5	12.0	413	12	BM059608	BM059608 KS01002E0
39	222.5	11.9	544	13	BQ591926	BQ591926 E012583-0
40	222.5	11.9	731	10	BG124857	BG124857 EST470503
41	222	11.9	640	9	AW738440	AW738440 EST339867
42	218.5	11.7	770	14	CB662235	CB662235 OSJNEd06B
43	216	11.6	448	9	AV644003	AV644003 AV644003
44	216	11.6	482	9	AV642898	AV642898 AV642898
45	216	11.6	502	9	AV626792	AV626792 AV626792

ALIGNMENTS

RESULT 1
BZ575759
LOCUS BZ575759 1323 bp DNA linear GSS 17-DEC-2002
DEFINITION msh2_4614.x1 msh Pseudomonas aeruginosa genomic clone msh2_4614, genomic survey sequence.
ACCESSION BZ575759
VERSION BZ575759.1 GI:27210820
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 1323)

```
  AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
  TITLE     Whole-Genome-Sequence variation among multiple isolates of
            Psedomonas aeruginosa library
  JOURNAL   J. Bacteriol., (2002) In press
COMMENT     Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .1323
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="MSH"
                     /db_xref="taxon:287"
                     /clone="msh2_4614"
                     /clone_lib="msh"
                     /note="Environmental isolate. Whole genomic shotgun
                     library."
BASE COUNT      240 a    394 c    318 g    371 t
ORIGIN

Alignment Scores:
Pred. No.:                9.61e-84       Length:          1323
Score:                    820.50         Matches:         171
Percent Similarity:       81.67%         Conservative:    34
Best Local Similarity:    68.13%         Mismatches:      37
Query Match:              43.97%         Indels:          12
DB:                       29             Gaps:            2

US-09-921-992-78 (1-372) x BZ575759 (1-1323)

Qy        115 GluGluArgIleArgMetValValAspCysAlaArgAspLysAsnIleProIleArgIle 134
              :::   |||::::::   |||||||||   ||||||:::::|||||||||||||||||||
Db         86 CAGCCCCGGGTCAAGGCCGTGGTCGACGCCGCGCGCGAGCGCAACATCCCGATCCGTATC 145

Qy        135 GlyValAsnAlaGlySerLeuGluLysAspLeuGlnGluLysTyrGlyGluProThrPro 154
              ||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db        146 GGCGTCAATGCCGGTTCGCTGGAAAAGGACCTGCAGAAGAAATACGGCGAACCGACCCCG 205

Qy        155 GlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeuAspArgLeuAsnPheAsp 174
              :::||||||||||||||||||||||||||||||||||||||||||:::|||:::|||
Db        206 GAAGCCCTGCTCGAATCGGCCATGCGCCACGTCGATCATCTCGACAAGCTGGACTTCCAG 265

Qy        175 GlnPheLysValSerValLysAlaSerAspValPheLeuAlaValGluSerTyrArgLeu 194
              |||||||||||||||||||||||||||||||:::||||||   :::|||||||||
Db        266 AACTTCAAGGTCAGCGTCAAGGCCTCCGACGTCTTCATGGCCGTCGCCGCCTATCGCCTG 325

Qy        195 LeuAlaLysGlnIleAspGlnProLeuHisLeuGlyIleThrGluAlaGlyGlyAlaArg 214
              ||||||:::||||||:::||||||||||||||||||||||||||||||||||   |||
Db        326 CTGGCCAGGCAGATCGAGCAGCCCCTGCACCTAGGCATCACCGAGGCCGGCGGCCTGCGC 385

Qy        215 SerGlyAlaValLysSerAlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAsp 234
              ||||||   ||||||||||:::||||||||:::||||||:::|||||||||||||||
Db        386 TCCGGCACGGTGAAGTCGGCGGTGGGGCTGGGCATGCTCCTGGCCGAGGGAATCGGCGAC 445

Qy        235 ThrLeuArgValSerLeuAlaAlaAspProValGluGluIleLysValGlyPheAspIle 254
              |||:::|||:::||||||||||||||||||||||||||||||||||||||||||||||
Db        446 ACCATCCGGATTTCCCTGGCTGCCGATCCGGTCGAGGAGATCAAGGTCGGTTTCGACATC 505

Qy        255 LeuLysSerLeuArgIleArgSerArgGlyIleAsnPheIleAlaCysProThrCysSer 274
              ||||||||||||   :::|||||||||||||||||||||||||||||||:::||||||
Db        506 CTCAAGTCCCTGCACCTGCGCTCCCGTGGCATCAACTTCATCGCCTGCCCGAGCTGTTCG 565

Qy        275 ArgGlnGluPheAspValIleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIle 294
              ||||||   |||||||||:::   |||:::|||   ||||||   ||||||||||||:::
Db        566 CGGCAGAACTTCGACGTGGTGAAGACCATGAACGAGCTGGAAGGGCGCCTGGAGGACCTG 625

Qy        295 IleThrProMetAspValSerIleIleGlyCysValValAsnGlyProGlyGluAlaLeu 314
```

```
              :::   ||||||||||||:::::|||   |||||||||||||||||||||||||
Db        626 CTGGTGCCGATGGACGTGGCCGTAATCGT-TGCGTCGTTAACGGTCCGGGCGAAGCCAAG 684

Qy        315 ValSerThrLeuGlyValThrGlyGlyAsnLysLysSerGlyLeuTyrGluAspGlyVal 334
                    :::|||   |||||||||      :::   |||:::|||      |||
Db        685 GAGCCCATGGTTGGCTC-ACCGGCGGCCC-TCCGAACCTGGTGTATATTCGACGGGAAAG 742

Qy        335 Arg---LysAspArgLeuAspAsnAsnAspMetIleAspGlnLeuGluAlaArgIleArg 353
                    :::                  ::::::|||||||||
Db        743 CCGGTGTCAGAATCTGGCCCAACGACAACCTTGGTGGACCAGCTT-------------- 787

Qy        354 AlaLysAlaSerGlnLeuAspGluAlaArgArg 364
                       :::   |||   |||:::|||
Db        788 ---------GTACGGTTTGATTCCGCCAAAAGG 811

RESULT 2
CB669733
LOCUS       CB669733                828 bp    mRNA    linear   EST 09-APR-2003
DEFINITION  OSJNEe02H05.f OSJNEe Oryza sativa (japonica cultivar-group) cDNA
            clone OSJNEe02H05 5', mRNA sequence.
ACCESSION   CB669733
VERSION     CB669733.1  GI:29673458
KEYWORDS    EST.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 828)
  AUTHORS   Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
            Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
  TITLE     Large-scale identification of ESTs involved in the interaction
            between rice and Magnaporthe grisea
  JOURNAL   Unpublished
COMMENT     Contact: Rod Wing
            Arizona Genomics Institute
            University of Arizona
            Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
            85721-0088, USA
            Tel: 520 626 3967
            Fax: 520 621 9288
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: gta aaa cga cgg cca gtg
            BACKWARD: gga aac agc tat gac cat g
            Plate: 02  row: H  column: 05
            Seq primer: gta aaa cga cgg cca gtg.
FEATURES             Location/Qualifiers
     source          1. .828
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /clone="OSJNEe02H05"
                     /tissue_type="Leaf"
                     /dev_stage="3 week"
                     /lab_host="DH10B"
                     /clone_lib="OSJNEe"
                     /note="Vector: pBluescript II KS +; Site_1: EcoRI; Site_2:
                     XhoI; 24 hrs after innoculation with Rice Blast (70-15)"
BASE COUNT      231 a    155 c    217 g    224 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:                3.95e-33       Length:          828
Score:                    379.00         Matches:         91
Percent Similarity:       55.27%         Conservative:    40
Best Local Similarity:    38.40%         Mismatches:      72
Query Match:              20.31%         Indels:          34
DB:                       14             Gaps:            6

US-09-921-992-78 (1-372) x CB669733 (1-828)
```

```
Qy         44 ThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArgValSerVal 63
              ||||||   ::::::      :::         |||||||||   ||||||::::::|||
Db          6 ACCGTAGAGGAGGTTATGAGGATAGCAGATAAAGGGGCTGATTTTGTTAGAATAACAGTC 65

Qy         64 ProThrMetAspAlaAlaGluAlaPhe--------------LysLeuIleLysGlnGln 78
                       |||:::|||                        |||::::::::::
Db         66 CAGGGTAGAAAGGAAGCTGATGCCTGCTTTGAGATTAAGAACACTCTTGTTCAGAAGAAT 125

Qy         79 ValAsnValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAla 98
              |||:::||||||||||||||||||||||||        :::||||||:::||||||
Db        126 TACAACATCCCCCTAGTGGCTGATATTCATTTTGCCCCGACAGTTGCTTTAAGAGTGGCT 185

Qy         99 GluTyrGlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGlu--------- 115
              |||         |||   :::|||:::||||||||||||      :::
Db        186 GAATGC---TTTGACAAAATTCGTGTCAACCCAGGGAATTTTGCTGATCGCCGTGCCCAA 242

Qy        116 ----------------------------------------------GluArgIleArgMet 120
                                                           |||   |||
Db        243 TTTGAGCAGCTTGAATATACTGAAGATGATTATCAAAAAGAGCTTGAGCATATCGAGAAG 302

Qy        121 Val------------ValAspCysAlaArgAspLysAsnIleProIleArgIleGlyVal 136
              |||            |||:::      :::                 :::||||||||||
Db        303 GTCTTCTCCCCGTTGGTTGAGAAATGCAAGCAGTATGGAAGAGCAATGCGTATAGGAACA 362

Qy        137 AsnAlaGlySerLeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAla 156
              |||   |||||||||         :::         ||||||:::   :::|||:::
Db        363 AATCATGGAAGTCTGTCTGACCGCATAATGAGTTACTATGGTGAT---TCTCCACGCGGA 419

Qy        157 LeuLeuGluSerAlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPhe 176
              ::::::||||||||||:::                       :::|||:::|||      |||
Db        420 ATGGTTGAGTCTGCTTTGGAATTTGCCAGGATCTGTCGGAAGCTGGACTTCCATAACTTT 479

Qy        177 LysValSerValLysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAla 196
                 |||:::||||||||||:::      :::   |||:::::::||||||||||||
Db        480 GTGTTTTCAATGAAAGCAAGTAACCCTGTTATCATGGTCCAAGCATATCGCTTGCTTGTA 539

Qy        197 LysGlnIle--------------AspGlnProLeuHisLeuGlyIleThrGluAlaGly 211
                 ::::::              |||   |||||||||||||||:::||||||||||||
Db        540 GCAGAAATGTATAACCTAGGGTGGGATTATCCTTTGCACTTGGGAGTTACAGAAGCTGGA 599

Qy        212 GlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeuLeuLeuSerGluGly 231
                       |||   :::|||||||||||||||||:::|||   ||||||   :::|||
Db        600 GAGGGTGAAGATGGGAGGATGAAGTCTGCCATTGGCATTGGAACACTTCTGATGGATGGC 659

Qy        232 IleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGluGluIle 248
              :::||||||||||:::||||||||||||||        |||   |||||||||
Db        660 TTGGGCGATACAATCCGTGTCTCCCTCACGGAACCACCTGAAGAAGAGATT 710

RESULT 3
CB619369
LOCUS       CB619369                853 bp    mRNA    linear   EST 08-APR-2003
DEFINITION  OSIIEa03O03.f OSIIEa Oryza sativa (indica cultivar-group) cDNA
            clone OSIIEa03O03 5', mRNA sequence.
ACCESSION   CB619369
VERSION     CB619369.1  GI:29614356
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 853)
  AUTHORS   Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
            Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
  TITLE     Large-scale identification of ESTs involved in the interaction
            between rice and Magnaporthe grisea
  JOURNAL   Unpublished
COMMENT     Contact: Rod Wing
            Arizona Genomics Institute
            University of Arizona
            Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
            85721-0088, USA
            Tel: 520 626 3967
            Fax: 520 621 9288
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: gta aaa cga cgg cca gtg
            BACKWARD: gga aac agc tat gac cat g
            Plate: 03  row: O  column: 03
            Seq primer: gta aaa cga cgg cca gtg.
FEATURES              Location/Qualifiers
     source           1. .853
                      /organism="Oryza sativa (indica cultivar-group)"
                      /mol_type="mRNA"
                      /cultivar="IR36"
                      /db_xref="taxon:39946"
                      /clone="OSIIEa03O03"
                      /tissue_type="Leaf"
                      /dev_stage="3 week"
                      /lab_host="DH10B"
                      /clone_lib="OSIIEa"
                      /note="Vector: pBluescript II KS +; Site_1: EcoRI; Site_2:
                      XhoI; Lesion Mimic SPL 11"
BASE COUNT      222 a    168 c    223 g    240 t
ORIGIN

Alignment Scores:
Pred. No.:                1.35e-29      Length:          853
Score:                    348.50        Matches:         81
Percent Similarity:       56.93%        Conservative:    34
Best Local Similarity:    40.10%        Mismatches:      58
Query Match:              18.68%        Indels:          29
DB:                       14            Gaps:            5

US-09-921-992-78 (1-372) x CB619369 (1-853)

Qy         74 LeuIleLysGlnGlnValAsnValProLeuValAlaAspIleHisPheAspTyrArgIle 93
              |||::::::::::      |||:::||||||||||||||||||||||||||||      :::
Db          8 CTTGTTCAGAAGAATTACAACATCCCCCTAGTGGCTGATATTCATTTTGCCCCGACAGTT 67

Qy         94 AlaLeuLysValAlaGluTyrGlyValAspCysLeuArgIleAsnProGlyAsnIleGly 113
              ||||||:::|||||||||         |||   :::|||:::||||||||||||
Db         68 GCTTTAAGAGTGGCTGAATGC---TTTGACAAAATTCGTGTCAACCCAGGGAATTTTGCT 124

Qy        114 AsnGlu---------------------------------------------------- 115
              :::
Db        125 GATCGCCGTGCCCAATTTGAGCAGCTTGAATATACTGAAGATGATTATCAAAAAGAGCTT 184

Qy        116 GluArgIleArgMetVal------------ValAspCysAlaArgAspLysAsnIlePro 131
              |||   |||       |||            |||:::      :::
Db        185 GAGCATATCGAGAAGGTCTTCTCCCCGTTGGTTGAGAAATGCAAGCAGTATGGAAGAGCA 244

Qy        132 IleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGlnGluLysTyrGlyGlu 151
              :::||||||||   |||   |||||||||          :::         ||||||:::
Db        245 ATGCGTATAGGAACAAATCATGGAAGTCTGTCTGACCGCATAATGAGTTACTATGGTGAT 304

Qy        152 ProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeuAspArgLeu 171
                 :::|||:::   ::::::||||||||||:::                      :::|||
Db        305 ---TCTCCACGCGGAATGGTTGAGTCTGCTTTGGAATTTGCCAGGATCTGTCGGAAGCTG 361

Qy        172 AsnPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAlaValGluSer 191
              :::|||       |||       |||:::||||||||||:::      :::   |||::::::
Db        362 GACTTCCATAACTTTGTGTTTTCAATGAAAGCAAGTAACCCTGTTATCATGGTCCAAGCA 421

Qy        192 TyrArgLeuLeuAlaLysGlnIle--------------AspGlnProLeuHisLeuGly 206
              ||||||||||||      ::::::                |||   ||||||||||||||||
Db        422 TATCGCTTGCTTGTAGCAGAAATGTATAACCTAGGGTGGGATTATCCTTTGCACTTGGGA 481

Qy        207 IleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeu 226
              :::||||||||||||          |||   :::|||||||||||||||:::|||
Db        482 GTTACAGAAGCTGGAGAGGGTGAAGATGGGAGGATGAAGTCTGCCATTGGCATTGGAACA 541

Qy        227 LeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGlu 246
              ||||||   :::|||:::||||||||||:::||||||||||||||         |||   |||
Db        542 CTTCTGATGGATGGCTTGGGCGATACAATCCGTGTCTCCCTCACGGAACCACCTGAAGAA 601
```

```
Qy        247 GluIle 248
              ||||||
Db        602 GAGATT 607

RESULT 4
AY104363
LOCUS       AY104363               1442 bp    mRNA    linear   HTC 16-OCT-2002
DEFINITION  Zea mays PCO116062 mRNA sequence.
ACCESSION   AY104363
VERSION     AY104363.1  GI:21207441
KEYWORDS    HTC.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 1442)
  AUTHORS   Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
            Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
  TITLE     Maize Mapping Project/DuPont Consensus Sequences for Design of
            Overgo Probes
  JOURNAL   Unpublished (2002)
REFERENCE   2  (bases 1 to 1442)
  AUTHORS   Coe,E.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-APR-2002) Maize Mapping Project, University of
            Missouri, Columbia, MO 65211, USA
COMMENT     If you are interested in getting corresponding physical clones,
            these are publicly available from ZmDB and may be found by BLAST
            searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR,
            www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the
            maize cDNA sequences is either Virginia Walbot, Stanford or Pat
            Schnable, Iowa State, then clones may be requested from ZmDB:
            www.zmdb.iastate.edu.
FEATURES             Location/Qualifiers
     source          1. .1442
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /db_xref="MaizeDB:636753"
                     /db_xref="taxon:4577"
                     /clone_lib="Maize Mapping Project/DuPont Cornsensus
                     Library"
                     /note="this sequence is part of a project of EST
                     assemblies resulting from the application of public
                     contigs to seed DuPont contigs; this resource was
                     assembled by DuPont as part of a collaboration for the
                     overgo addressing of BACs in conjunction with the Maize
                     Mapping Project"
BASE COUNT      381 a    282 c    374 g    399 t      6 others
ORIGIN

Alignment Scores:
Pred. No.:              9.68e-27        Length:         1442
Score:                  326.50          Matches:        99
Percent Similarity:     49.66%          Conservative:   46
Best Local Similarity:  33.90%          Mismatches:     77
Query Match:            17.50%          Indels:         71
DB:                     11              Gaps:           11

US-09-921-992-78 (1-372) x AY104363 (1-1442)

Qy         65 ThrMetAspAlaAlaGluAlaPheLys-----------LeuIleLysGlnGlnValAsn 80
              |||   :::|||         |||:::           |||::::::::::     |||
Db          3 ACGAGGGAAGCTGATGCCTGCTTTGAGATCAAGAACACTCTGGTTCAGAAGAATTACAAC 62

Qy         81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100
              :::||||||||||||||||||||||||        :::|||||::|||||||||
Db         63 ATTCCACTAGTGGCCGATATTCATTTTGCTCCTACGGTAGCCCTAAGGGTGGCAGAATGT 122

Qy        101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGlu-------------- 115
                  |||   :::|||:::|||||||||||||      :::
Db        123 ---TTTGACAAAATTCGTGTGAACCCAGGAAATTTTGCTGATCGTCGTGCTCAATTTGAA 179

Qy        116 ------------------------------------GluArgIleArgMetVal--- 121
                                                  |||   |||      |||
Db        180 AAGCTGGAATATACTGACGACGACTACCAAAAAGAGCTAGAGCATATTGAGAAGGTGTTT 239

Qy        122 ---------ValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAla 138
                      |||:::     :::               :::||||||||||   |||
Db        240 TCTCCATTAGTTGAGAAATGCAAGCAGTATGGAAGAGCAATGCGAATAGGAACAAATCAT 299

Qy        139 GlySerLeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeu 158
              |||||||||          :::         ||||||:::   :::|||:::   ::::::
Db        300 GGTAGTCTTTCTGACCGCATAATGAGCTACTATGGTGAT---TCTCCACGGGGAATGGTT 356

Qy        159 GluSerAlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysVal 178
              |||||||||:::                    :::|||:::|||      |||
Db        357 GAGTCTGCTTTGGAATTTGCTAGAATATGTCGGAAGTTGGACTTCCACAACTTTGTATTT 416

Qy        179 SerValLysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGln 198
              |||:::||||||||||:::      :::   |||:::::::||||||||||||      :::
Db        417 TCGATGAAAGCTAGTAACCCTGTTATCATGGTCCAAGCATATCGCCTGCTTGTGGCAGAA 476

Qy        199 Ile---------------AspGlnProLeuHisLeuGlyIleThrGluAlaGlyGlyAla 213
              :::               |||   ||||||||||||||:::||||||||||||
Db        477 ATGTATAACCTAGGATGGGATTATCCCTTGCACTTGGGTGTTACAGAAGCTGGAGAGGGT 536

Qy        214 ArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGly 233
                  :::|||||||||||||||:::|||   ||||||   :::|||:::|||
Db        537 GAAGATGA-AGGATGAAATCTGCTATTGGCATTGGGACACTGCTAATGGATGGTTTGGGT 595

Qy        234 AspThrLeuArgValSerLeuAlaAlaAspProValGluGluIle--------------- 248
              |||||||:::|||||||||||||         |||   |||||||||
Db        596 GATACAATCCGTGTCTCCCTCACAGAACCACCAGAAGAAGAGATTGATCCTTGCCAAAGG 655

Qy        249 ---------------------------------LysValGlyPheAspIleLeuLysSe 257
                                                |||:::|||:::      ||||||
Db        656 TTGGCAAATCTTGGGACGCAGGCCGCAAACCTTCAAATTGGGGTGGCCCCATTTGAAG-- 713

Qy        257 rLeuArgIleArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGl 277
                  :::|||   |||:::      |||:::
Db        714 ----------AAAAGCACAGGCGTTATTTTGATTTCCAGC-------------------- 743

Qy        277 uPheAspValIleGlyThrValAsnAlaLeu---GluGlnArgLeuGluAspIleIleTh 296
                  :::|||   ||||||   |||      :::|||:::      :::   ||
Db        744 ----------GTAGGAGTGGTCAATTGCCTTTGCAGAAGGAGGGTGAGGAAGTTGACTAC 793

Qy        296 rProMetAspValSerIleIleGlyCysValVal 307
              |   |||              |||:::|||
Db        794 AGAAATG---------------TCCTGCATCGTG 812

RESULT 5
BZ569685/c
LOCUS       BZ569685               1268 bp    DNA     linear   GSS 17-DEC-2002
DEFINITION  pacs2-164_970.s1 pacs2-164 Pseudomonas aeruginosa genomic clone
            pacs2-164_970, genomic survey sequence.
ACCESSION   BZ569685
VERSION     BZ569685.1  GI:27204646
KEYWORDS    GSS.
SOURCE      Pseudomonas aeruginosa
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 1268)
  AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
  TITLE     Whole-Genome-Sequence variation among multiple isolates of
            Psedomonas aeruginosa library
  JOURNAL   J. Bacteriol.,  (2002) In press
COMMENT     Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
```

```
          Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .1268
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="2-164"
                     /db_xref="taxon:287"
                     /clone="pacs2-164_970"
                     /clone_lib="pacs2-164"
                     /note="clinical isolate 2-164 Whole genomic shotgun
                     library."
BASE COUNT      257 a    386 c    351 g    273 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:              3.05e-26        Length:         1268
Score:                  321.50          Matches:        119
Percent Similarity:     62.55%          Conservative:   33
Best Local Similarity:  48.97%          Mismatches:     60
Query Match:            17.23%          Indels:         31
DB:                     29              Gaps:           8

US-09-921-992-78 (1-372) x BZ569685 (1-1268)

Qy        142 GluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSerAla 161
              ||||||   |||   ||||||:::         |||||||:::
Db       1195 GAAAAGGCTCTA---GAAAAATTTCGGAACCGGACCCCAAAG---------------- 1157

Qy        162 MetArgHisValAspHisLeuAspArgLeuAsn-----------PheAspGlnPheLys 177
              :::         |||         ||||||              |||         |||
Db       1156 CTTTGTGAATGGGATATGGCAATTCGGTTATTTTGAGAAAGTTGATTCCAAGACTCTAAG 1097

Qy        178 ValSerValLysAlaSerAspValPhe-LeuAlaVal-GluSerTyrArgLeuLeuAlaL 197
              |||   |||:::||||||   |||||| :::|||:::    :::||||||||||||   |
Db       1096 GTCGGTGTCCAGGCTTCAAAAGTGTTTAATGGCCTTGACGCGCTTACCGCTTGTTGGGAA 1037

Qy        197 ysGlnIle---AspGlnProLeuHisLeuGlyIleThrGluAlaGly------GlyAlaA 214
              ||||||||       ||||||:::|||||||||   |||   |||      |||   |
Db       1036 AGCAAATTGGAGGAAACCCCTTGAACTTGGGGATCAACGAAGGCGGGCGGCCTGGGTTTC 977

Qy        214 rgSerGlyAlaValLysSerAlaIleGlyLeu-GlyLeuLeuLeuSerGluGlyIle--- 232
              ||         :::   ||||||:::|||||| |||:::   |||      ||||||
Db        976 GGGAA---CGTTTAAACTCGGCGTTGGGGCTGGGGAATTTTCCTTGGCCGAGGGATTCGG 920

Qy        233 GlyAspThrLeuArgValSerLeuAlaAla--AspProValGlu-GluIleLysValGly 251
                  |||:::|||||||||     ||||||||||| ||||||:::||||||
Db        919 CGAAAACNATCCCGGATTTCCCTGGCTTGCCGGATCCGGTTGAAGGAGATCCAGGTCGGT 860

Qy        252 Phe-AspIleLeuLys-SerLeuArg-IleArgSerArgGlyIle-AsnPheIleAlaCy 270
              ||| ||||||||||| ||||||    :::|||||||||||||| |||   ||||||||
Db        859 TTCGGACATCCTTAAGTTCCCTGAACCTTGCGCTCCCGTGGCATAAAACTCCATTGCTTG 800

Qy        270 sPro-ThrCysSer-ArgGlnGlu-PheAspValIleGlyThrVal-AsnAlaLeuGluG 289
              |||| :::|||||| |||||||| |||||||||:::   |||::: |||   ||||||
Db        799 CCCGGAGCTGTTCGGCGGCAAGAACTTCGACGTGTTGAAGACCATGAAACGAGCTGGAAG 740

Qy        289 lnArgLeuGluAspIleIleThrProMetAspValSerIleIleGlyCysValValAsnG 309
              ||||||||||||:::::   ||||||||||||:::::|||||||||||||||||||||
Db        739 GGCGCCTGGAGGACCTGCTGATGCCGATGGACGTGGCCGTGATCGGTTGCGTCGTCAACG 680

Qy        309 lyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyGlyAsnLysLysSerGlyL 329
              |||||   ||||||      :::   :::|||:::||||||                 :
Db        679 GTCCGTGCGAAGCCAAGGAGGCCCATGTCGGCCTCACCGGCTGCACTCCGAACCTG---G 623

Qy        329 euTyrGluAspGlyValArgLysAspArgLeuAspAsnAsnAspMetIleAspGlnLeuG 349
              ::|||   ||||||            :::|||   |||:::::::::::::|||:::||||
Db        622 TGTATATCGACGGCAAGCCGTCGCAGAAACTGACCAACGACAACCTGGTGGACGAGCTGG 563

Qy        349 luAlaArgIleArgAlaLysAlaSerGlnLeuAspGluAla 362
              ||      ||||||   ||||||::::::      ||||||
Db        562 AACGGCTGATCCGCCAGAAGGCGGCCGAGAAGGCCGAGGCC 522

RESULT 6
```

```
BI920888
LOCUS       BI920888                579 bp    mRNA    linear   EST 10-MAR-2003
DEFINITION  EST540823 potato microtubers, in vitro-grown Solanum tuberosum cDNA
            clone cSTE26D20 5' end, mRNA sequence.
ACCESSION   BI920888
VERSION     BI920888.1  GI:16216916
KEYWORDS    EST.
SOURCE      Solanum tuberosum (potato)
  ORGANISM  Solanum tuberosum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; lamiids; Solanales; Solanaceae; Solanum.
REFERENCE   1  (bases 1 to 579)
  AUTHORS   van der Hoeven,R., Bezzerides,J., Bachem,C., Visser,R., Karamycheva
            ,S.A., Tsai,J., Van Aken,S., Utterback,T., Chiemingo,A., Bougri,O.,
            Buell,C.R., Ronning,C., Tanksley,S. and Baker,B.
  TITLE     Generation of ESTs from in vitro grown microtubers (2001b)
  JOURNAL   Unpublished
COMMENT     Contact: Robin Buell
            The Institute for Genomic Research
            9712 Medical Center Dr, Rockville, MD 20850, USA
            Email: potato-array@tigr.org
            This clone can be obtained from the University of Arizona Genomics
            Institute. Orders can be made through URL:
            http://genome.arizona.edu/orders/
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .579
                     /organism="Solanum tuberosum"
                     /mol_type="mRNA"
                     /cultivar="Bintje"
                     /db_xref="taxon:4113"
                     /clone="cSTE26D20"
                     /tissue_type="axillary buds of stem explants; growing
                     sink-tubers"
                     /dev_stage="7, 8 and 10 days"
                     /lab_host="SOLR"
                     /clone_lib="potato microtubers, in vitro-grown"
                     /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
                     XhoI; Tissue supplied by Christian Bachem and Richard
                     Visser (Department of Plant Breeding, Wageningen
                     University, The Netherlands); sequencing by The Institute
                     for Genomic Research. The cSTA libraries will attempt to
                     capture the induction and initiation/initial growth of the
                     tuber in an in vitro system as described in Bachem et al.
                     (Plant Journal, 1996). Small microtubers develop from
                     axillary buds attached to stem explants when placed on a
                     high sucrose medium (10%). Visible morphological changes
                     occur synchronously at day five in the axillary buds. The
                     first library, cSTA (1-20) consists of axillary buds
                     harvested on days 1-3. This targets those genes involved
                     in induction of the microtubers. The following libraries,
                     cSTA (21-40) and cSTA (41-60), capture genes involved in
                     tuber initiation and outgrowth. This library is noted as
                     P3  in Tanksley lab notebooks."
BASE COUNT      157 a    106 c    153 g    163 t
ORIGIN

Alignment Scores:
Pred. No.:              5.01e-26        Length:         579
Score:                  315.50          Matches:        74
Percent Similarity:     55.85%          Conservative:   31
Best Local Similarity:  39.36%          Mismatches:     54
Query Match:            16.91%          Indels:         29
DB:                     12              Gaps:           4

US-09-921-992-78 (1-372) x BI920888 (1-579)

Qy         80 AsnValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGlu 99
              |||:::|||||||||||||||||||||||||         :::||||||:::|||||||||
Db         11 AACATCCCTCTGGTGGCTGACATTCATTTTGCTCCTTCTGTTGCACTTCGAGTGGCGGAG 70

Qy        100 TyrGlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsn---------------- 114
```

```
                   |||   ::|||:::|||||||||||        :::
Db         71 TGC---TTTGACAAAATACGTGTCAATCCTGGAAACTTTGCTGACAGGCGAGCCCAGTTT 127

Qy        115 ---------------------------------------------------GluGluArg 117
                                                                 ||||||
Db        128 GAGCAATTAGAGTACACAGAAGATGACTATCAGAAAGAACTCGAGCATATTGAGGAGGTT 187

Qy        118 IleArgMetValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsn 137
                           :::|||::      :::                 :::||||||||| |||
Db        188 TTTACACCATTGGTGGAAAAATGTAAGAAGTATGGACGTGCAATGCGCATTGGGACAAAC 247

Qy        138 AlaGlySerLeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeu 157
                 ||||||         :::             ||||||:::   :::|||:::   :::
Db        248 CATGGGAGCCCTTCAGATCGCATTATGAGCTATTATGGGGAC---TCGCCTAGGGGAATG 304

Qy        158 LeuGluSerAlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLys 177
              :::|||||||||                             :::|||:::|||      |||
Db        305 GTAGAATCAGCATTTGAGTTTGCAAGAATTTGTAGAAAGTTGGACTTTCACAATTTCGTC 364

Qy        178 ValSerValLysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLys 197
              |||:::||||||||||:::       :::   |||::::::||||||||||||
Db        365 TTCTCAATGAAAGCTAGCAATCCAGTAGTTATGGTTCAGGCGTATCGCCTTCTTCTAGCT 424

Qy        198 GlnIle---------------AspGlnProLeuHisLeuGlyIleThrGluAlaGlyGly 212
              ::::::                |||   ||||||||||||||||:::|||||||||||||
Db        425 GAGATGTATGTTCAGGGATGGGACTATCCTTTACACTTGGGAGTTACTGAAGCTGGTGAA 484

Qy        213 AlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIle 232
                  |||   :::||||||||||||||:::|||   ||||||   :::|||:::
Db        485 GGTGAGGATGGAAGGATGAAGTCTGCAATCGGTATTGGCACACTTCTTCAGGATGGTTTG 544

Qy        233 GlyAspThrLeuArgValSerLeu 240
              |||||||||:::|||||||||||
Db        545 GGTGATACAATACGGGTTTCCCTA 568
```

RESULT 7
BH235010

```
LOCUS       BH235010                 720 bp    DNA     linear   GSS 01-JAN-2002
DEFINITION  PH_05.x Spiroplasma kunkelii H Spiroplasma kunkelii genomic clone
            PH_05.x, genomic survey sequence.
ACCESSION   BH235010
VERSION     BH235010.1  GI:18030478
KEYWORDS    GSS.
SOURCE      Spiroplasma kunkelii
  ORGANISM  Spiroplasma kunkelii
            Bacteria; Firmicutes; Mollicutes; Entomoplasmatales;
            Spiroplasmataceae; Spiroplasma.
REFERENCE   1  (bases 1 to 720)
  AUTHORS   Hogenhout,S.A.
  TITLE     Genomic sequences from Spiroplasma kunkelii strain M2
  JOURNAL   Unpublished
COMMENT     Contact: Hogenhout SA
            Department of Entomology
            The Ohio State University-OARDC
            120 Thorne Hall, 1680 Madison Avenue, Wooster, OH 44691, USA
            Tel: 330 263 3730
            Fax: 330 263 3686
            Email: hogenhout.1@osu.edu
            Class: HindIII fragments.
FEATURES             Location/Qualifiers
     source          1..720
                     /organism="Spiroplasma kunkelii"
                     /mol_type="genomic DNA"
                     /strain="M2"
                     /db_xref="taxon:47834"
                     /clone="PH_05.x"
                     /clone_lib="Spiroplasma kunkelii H"
BASE COUNT      255 a     96 c    132 g    233 t      4 others
ORIGIN

Alignment Scores:
Pred. No.:              1.32e-25      Length:        720
Score:                  313.00        Matches:       67
Percent Similarity:     62.20%        Conservative:  35
Best Local Similarity:  40.85%        Mismatches:    61
Query Match:            16.77%        Indels:        2
DB:                     28            Gaps:          0
```

US-09-921-992-78 (1-372) x BH235010 (1-720)

```
Qy        190 GluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeuGlyIleThrGlu 209
              :::::|||   |||   :::|||:::   :::   |||||||||||||||:::   |||
Db         31 CAAGCTTATACCTTAGCTAGTAAAGAATGAAATTATCCCCTTCATCTAGGGATGCT-GAA 89

Qy        210 AlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeuLeuLeuSer 229
              ||||||         :::|||      ||||||      ||||||      ||||||
Db         90 GCTGGTAGTCATCATACCGGAACAANNAAATCATGTAGTGGTTTGTCACCGCTTCTTTTT 149

Qy        230 GluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGluGluIleLys 249
                 ||||||||||||||:::|||:::||||||:::   ||||||:::   |||::::::
Db        150 AATGGTATTGGTGATACGATCCGAATTAGTTTATCAACTGATCCAATAGCAGAGGTGGAA 209

Qy        250 ValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsnPheIleAla 269
              ||||||      :::|||   ||||||            ::::::   :::|||::
Db        210 GTTGGTAAACGAATGTTAAATTCATTAGGACTTTATGATAACATTGTTGATATAATTGCT 269

Qy        270 CysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsnAlaLeuGluGln 289
              ||||||||||||   |||   |||:::|||:::            |||   ||||||:::
Db        270 TGCCCAACATGTGGTCGTTTAGAATATGACCTTTTCCCCGTTGTAAAAGAATTAGAAGAA 329

Qy        290 ArgLeuGluAspIleIleThrProMetAspValSerIleIleGlyCysValValAsnGly 309
                 :::|||:::         |||:::   ::::::|||:::||||||||||||||||||
Db        330 TATACAAAAGATTTAAAAATTCCATTAAAAATTGCCATTTTAGGTTGTGTTGTTAATGGT 389

Qy        310 -ProGlyGluAlaLeuValSerThrLeuGlyValThrGlyGlyAsnLysLysSerGlyLe 329
                ||||||||||||      :::   ||||||:::   ||||||                ::
Db        390 CCCAGGAGAAGCTAAACAAGCTGATTTAGGGATTGCTGGTGGTAAAAATGGTGGAATTAT 449

Qy        329 uTyrGluAspGlyValArgLysAspArgLeuAspAsnAsnAspMetIleAspGlnLeuGl 349
              :::::::   |||                                |||:::::   :::|||::
Db        450 TTTTAAAAAAGGTAAAATTTATAAATCATGTAAACAAGAAGATCTTGTTCCAGAATTAAA 509

Qy        349 uAlaArgIle 352
                :   |||
Db        510 ACTATTAATT 519
```

RESULT 8
AZ570993/c

```
LOCUS       AZ570993                 658 bp    DNA     linear   GSS 15-MAY-2001
DEFINITION  281PvD07 Pv MBN #30 Plasmodium vivax genomic 3', genomic survey
            sequence.
ACCESSION   AZ570993
VERSION     AZ570993.1  GI:13982637
KEYWORDS    GSS.
SOURCE      Plasmodium vivax (malaria parasite P. vivax)
  ORGANISM  Plasmodium vivax
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 658)
  AUTHORS   Carlton,J.M.-R. and Dame,J.B.
  TITLE     The Plasmodium vivax and P. berghei gene sequence tag projects
  JOURNAL   Parasitol. Today (Regul. Ed.) 16 (10), 409 (2000)
COMMENT     Contact: Dame JB
            Dept. of Pathobiology, College of Veterinary Medicine
            University of Florida
            2015 SW 23rd Avenue, Bldg 1017, Gainesville, FL 32611, USA
            Tel: 352 392 4700
            Fax: 352 392 9704
            Email: damej@mail.vetmed.ufl.edu
            Seq primer: M13(-20) forward
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1..658
                     /organism="Plasmodium vivax"
                     /mol_type="genomic DNA"
                     /strain="Salvador I (Collins, W. 1972. J. Parasitol. 69,
                     497-598)"
```

/db_xref="taxon:5855"
/dev_stage="asexual blood forms"
/lab_host="Saimiri boliviensis"
/clone_lib="Pv MBN #30"
/note="Vector: pBluescript SK(+) vector DNA, phagemid excised from lambda ZAP; Site_1: EcoR V; Site_2: EcoR V; Host leukocytes were extracted from P. vivax infected blood using the following methods: first, infected blood was activated by the addition of 0.5 ml of ADP (40mg/ml) per 10 ml blood. Then blood was passed over a column of acid washed 0.1 mm glass beads, then through a Plasmodipur filter, followed by passage through a column of pre-wet Whatman CF11 powder (1:2 ratio volume of blood to CF11), and finally centrifuged through a 50% Percoll density cushion. Purified DNA was digested with mung bean nuclease in the presence of 44% formamide at 50oC as described (Vernick, K.D., Imberski, R.B., and McCutchan, T.F. 1988. Nucleic Acids Research 16:6883-6896). Digested DNA was blunt-ended using T4 DNA Polymerase and size fractionated over a Sepharose CL-2B column. Fractions in the size range 500bp-4kb were ligated into the Eco RV site of pBluescript SK(+), and E. coli XL-10 Gold transformed with the ligation mixture."

BASE COUNT 178 a 150 c 119 g 206 t 5 others
ORIGIN

Alignment Scores:

Pred. No.:	5.01e-25	Length:	658
Score:	307.50	Matches:	79
Percent Similarity:	53.85%	Conservative:	33
Best Local Similarity:	37.98%	Mismatches:	61
Query Match:	16.48%	Indels:	35
DB:	28	Gaps:	6

US-09-921-992-78 (1-372) x AZ570993 (1-658)

```
Qy         73 LysLeuIleLysGlnGlnValAsnValProLeuValAlaAspIleHisPheAspTyrArg 92
              |||:::      |||   :::      ||||||||||||||||||||          :::
Db        642 AAAATGGAAGCTCAAAATATTACATTTCCGTTAGTAGCAGACATACATTTAATCCCAAAA 583

Qy         93 IleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArgIleAsnProGlyAsnIle 112
              |||:::|||      |||:::         |||   :::||||||   |||||||||
Db        582 ATTTCTCTCATGCAAGCAGATGTA---TTTGATAAAATACGCATAACCCCAGGGAATTAC 526

Qy        113 GlyAsnGluGluArg---------------------------------------------- 117
                 :::      :::
Db        525 GTCGATGGGAGNAAGAAATGGATCAATAAGATTTACAAAACGAGAGAAGANTTCGATCAA 466

Qy        118 ----------------------------IleArgMetValValAspCysAlaArgAspLys 128
                                         |||   ::::::      |||   |||
Db        465 GGGAAATTATTTATTCAGNAGAAGTTCATCCCCTTGATTGAAAAGTGTAAGAGG---TTA 409

Qy        129 AsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGlnGluLys 148
              |||      ||||||||||||   |||   |||||||||        :::
Db        408 AATAGAGCCATACGAATCGGCACAAATCACGGCTCCCTATCTTNTAGGGTACTCTCTTTC 349

Qy        149 TyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeu 168
              ||||||:::   ||||||      ::::::|||||||||                |||
Db        348 TATGGGGAT---ACCCCNTTAGGTATGGTCGAATCGGCGTTTGAATTTTCCGATTTGTGT 292

Qy        169 AspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAla 188
                 :::   ||||||                |||:::||||||||||:::   ::::::
Db        291 GTACAGAATAACTTCTTCAATGTGGTTTTCTCCATGAAGGCTTCCAACGCGTATATAATG 232

Qy        189 ValGluSerTyrArgLeuLeu---AlaLysGlnIleAspGln----------------- 201
              ::::::|||||||||||||||   |||:::|||   ::::::
Db        231 ATACAGTCGTATAGGCTACTCGTGGCGAGGCAATATGAAAGGATGGGCAACGGGTTACTA 172

Qy        202 ---ProLeuHisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSer 220
                 ||||||||||||||:::|||||||||||||         :::|||   :::||||||
Db        171 TTCCCACTGCATTTGGGCGTTACCGAAGCAGGGTTTGGGGATAACGGAAGGATAAAATCC 112

Qy        221 AlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240
```

```
                  :::|||:::|||   ||||||   :::|||||||||||||||:::|||:::||||||
Db        111 TATTTGGGGATAGGTTCGCTGCTGTACGACGGGATAGGAGACACCATCAGGATTTCGTTA 52

Qy        241 AlaAlaAspProValGluGluIle 248
                    ||||||   |||||||::
Db         51 ACTGAGGACCCTTGGGAGGAATTA 28
```

RESULT 9
CB628479

LOCUS	CB628479 852 bp mRNA linear EST 08-APR-2003
DEFINITION	OSIIEb04B06.f OSIIEb Oryza sativa (indica cultivar-group) cDNA clone OSIIEb04B06 5', mRNA sequence.
ACCESSION	CB628479
VERSION	CB628479.1 GI:29623468
KEYWORDS	EST.
SOURCE	Oryza sativa (indica cultivar-group)
ORGANISM	Oryza sativa (indica cultivar-group) Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE	1 (bases 1 to 852)
AUTHORS	Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E., Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
TITLE	Large-scale identification of ESTs involved in the interaction between rice and Magnaporthe grisea
JOURNAL	Unpublished
COMMENT	Contact: Rod Wing Arizona Genomics Institute University of Arizona Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ 85721-0088, USA Tel: 520 626 3967 Fax: 520 621 9288 Email: http://genome.arizona.edu PCR PRimers FORWARD: gta aaa cga cgg cca gtg BACKWARD: gga aac agc tat gac cat g Plate: 04 row: B column: 06 Seq primer: gta aaa cga cgg cca gtg.

FEATURES Location/Qualifiers
source 1. .852
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="mRNA"
/cultivar="IR36"
/db_xref="taxon:39946"
/clone="OSIIEb04B06"
/tissue_type="Leaf"
/dev_stage="3 week"
/lab_host="DH10B"
/clone_lib="OSIIEb"
/note="Vector: pBluescript II KS +; Site_1: EcoRI; Site_2: XhoI; 24 hrs after innoculation with Rice Blast (PO6-6-3)"

BASE COUNT 233 a 172 c 223 g 224 t
ORIGIN

Alignment Scores:

Pred. No.:	2.69e-21	Length:	852
Score:	276.50	Matches:	72
Percent Similarity:	53.43%	Conservative:	37
Best Local Similarity:	35.29%	Mismatches:	66
Query Match:	14.82%	Indels:	30
DB:	14	Gaps:	5

US-09-921-992-78 (1-372) x CB628479 (1-852)

```
Qy          8 GlnArgArgLysSerThrArgIleTyrValGlyAsnValProIleGlyAspGlyAlaPro 27
              :::||||||||||:::      :::   |||||||||||||||:::|||          |||
Db        238 AGGAGGAGAAAAACCCGCACTGTGATGGTTGGGAATGTGCCACTTGGCAGTGATCATCCC 297

Qy         28 IleAlaValGlnSerMetThrAsnThrArgThrThrAspValGluAlaThrValAsnGln 47
              |||   :::|||:::||||||   :::   |||   ||||||      ||||||   :::
Db        298 ATTAGGATTCAGACTATGACCACCTCGGATACCAAGGATGTTGCTAAAACCGTAGAGGAG 357
```

```
Qy         48 IleLysAlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMetAsp 67
              :::       :::          |||||||||    ||||||::::::|||
Db        358 GTTATGAGGATAGCAGATAAAGGGGCTGATTTTGTTAGAATAACAGTCCAGGGTAGAAAG 417

Qy         68 AlaAlaGluAlaPhe--------------LysLeuIleLysGlnGlnValAsnValPro 82
              |||:::|||                    |||:::::::::      |||:::|||
Db        418 GAAGCTGATGCCTGCTTTGAGATTAAGAACACTCTTGTTCAGAAGAATTACAACATCCCC 477

Qy         83 LeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyVal 102
              ||||||||||||||||||||          :::||||||::|||||||||
Db        478 CTAGTGGCTGATATTCATTTTGCCCCGACAGTTGCTTTAAGAGTGGCTGAATGC---TTT 534

Qy        103 AspCysLeuArgIleAsnProGlyAsnIleGlyAsnGlu---------------------- 115
              |||   :::|||:::||||||||||||||      :::
Db        535 GACAAAATTCGTGTCAACCCAGGGAATTTTGCTGATCGCCGTGCCCAATTTGAGCAGCTT 594

Qy        116 --------------------------------GluArgIleArgMetVal--------- 121
                                             |||   |||      |||
Db        595 GAATATACTGAAGATGATTATCAAAAAGAGCTTGAGCATATCGAGAAGGTCTTCTCCCCG 654

Qy        122 ---ValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140
                 |||:::     :::                :::||||||||||   |||   ||||||
Db        655 TTGGTTGAGAAATGCAAGCAGTATGGAAGAGCAATGCGTATAGGAACAAATCATGGAAGT 714

Qy        141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSer 160
              |||       :::        |||||||:::   :::|||:::   :::::||||||
Db        715 CTGTCTGACCGCATAATGAGTTACTATGGTGAT---TCTCCACGCGGAATGGTTGAGTCT 771

Qy        161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180
              |||:::           :::   :::|||::||||      |||      |||:::
Db        772 GCTTTGGAATTTGCAGGATCTGTCGG-AAGCTGGACTTCCATAACTTTGTGTTTTCAATG 830

Qy        181 LysAlaSerAsp 184
              ||||||||||:::
Db        831 AAAGCAAGTAAC 842

RESULT 10
CA022320
LOCUS       CA022320                602 bp    mRNA    linear   EST 23-OCT-2002
DEFINITION  HZ42M11r HZ Hordeum vulgare subsp. vulgare cDNA clone HZ42M11
            5-PRIME, mRNA sequence.
ACCESSION   CA022320
VERSION     CA022320.1  GI:24299694
KEYWORDS    EST.
SOURCE      Hordeum vulgare subsp. vulgare
  ORGANISM  Hordeum vulgare subsp. vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 602)
  AUTHORS   Radchuk,V., Zhang,H., Weschke,W., Potokina,E. and Wobus,U.
  TITLE     Barley ESTs from developing seeds
  JOURNAL   Unpublished
COMMENT     Contact: Stein Nils
            Molecular Markers Group, Department Genbank
            Institute of Plant Genetics and Crop Plant Research (IPK)
            Corrensstr. 3, 06466, Gatersleben, Germany
            Tel: 039482-5522
            Fax: 039482-5595
            Email: stein@ipk-gatersleben.de
            Insert Length: 602   Std Error: 0.00
            Plate: 42  row: M  column: 11
            Seq primer: M13rev.
FEATURES             Location/Qualifiers
     source          1. .602
                     /organism="Hordeum vulgare subsp. vulgare"
                     /mol_type="mRNA"
                     /cultivar="barke"
                     /db_xref="GABI:275726"
                     /db_xref="taxon:112509"
                     /clone="HZ42M11"
                     /tissue_type="pericarp"
                     /dev_stage="0-7 DAP (days after pollination)"
                     /lab_host="XL10-Gold"
                     /clone_lib="HZ"
                     /note="Vector: pBluescript SK+; Site_1: EcoRI (5'-end of
                     cDNA); Site_2: XhoI (3'-end of cDNA); pericarp 0-7
                     DAP(days after pollination). Due to a cloning artefact
                     caused by the kit, in most cases the EcoRI site is NOT
                     present, as well as the EcoRIadapter used for cloning. To
                     excise the insert, restriction sites upstream EcoRI should
                     be used (e.g. BamHI, SalI,PstI). NOTE: Also due to the
                     cloning system used Blue/white selection for recombinats
                     is not 100% reliable.Average insert size is 900 bp"
BASE COUNT      169 a    110 c    156 g    167 t
ORIGIN

Alignment Scores:
Pred. No.:                  7.14e-21        Length:          602
Score:                      271.00          Matches:         70
Percent Similarity:         51.52%          Conservative:    32
Best Local Similarity:      35.35%          Mismatches:      62
Query Match:                14.52%          Indels:          34
DB:                         14              Gaps:            6

US-09-921-992-78 (1-372) x CA022320 (1-602)

Qy         55 GlyAlaAspIleValArgValSerValProThrMetAspAlaAlaGluAlaPhe------ 72
              |||||||||   ||||||::::::|||                 |||:::|||
Db         15 GGCGCTGATTTTGTTAGAATAACCGTCCAGGGTAAAAAGGAAGCTGATGCCTGCTTTGAG 74

Qy         73 ---------LysLeuIleLysGlnGlnValAsnValProLeuValAlaAspIleHisPhe 89
                      |||::::::::::      |||:::|||||   ||||||||||||||
Db         75 ATTAAGAACACTCTTGTCCAGAAGAATTACAACATCCCTCTAGCGGCCGATATCCATTTT 134

Qy         90 AspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArgIleAsnPro 109
                 :::||||||:::|||||||||          |||   :::|||:::||||||
Db        135 GCGCCTACAGTAGCTTTAAGAGTGGCTGAATGC---TTTGACAAAATCCGTGTTAACCCA 191

Qy        110 GlyAsnIleGlyAsnGlu------------------------------------------ 115
              ||||||     :::
Db        192 GGAAACTTTGCCGATCGCCGTGCCCAATTTGAAAAGCTGGAATATACTGAAGACGATTAC 251

Qy        116 ------------GluArgIleArgMetVal------------ValAspCysAlaArgAsp 127
                          |||   |||      |||            |||:::     :::
Db        252 GAAAAGGAGCTTGAACACATTGAGAGGGTCTTTTCTCCATTGGTTGAGAAATGCAAGAAG 311

Qy        128 LysAsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGlnGlu 147
                    :::|||||||||   |||   |||||||||         :::
Db        312 TATGGAAGAGCCATGCGTATCGGAACAAATCATGGTAGTCTTTCTGACCGGATAATGAGC 371

Qy        148 LysTyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHis 167
              |||||||:::   :::|||:::   ::::::||||||||||:::
Db        372 TACTATGGTGAT---TCTCCAAGGGGAATGGTTGAGTCTGCTTTGGAATTTGCTAGGATC 428

Qy        168 LeuAspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspValPheLeu 187
                  |||:::|||      |||      |||:::||||||||||:::         :::
Db        429 TGTCGGAATTTGGACTTCCATAACTTTGTATTTTCAATGAAAGCAAGTAACCCTGTTGTC 488

Qy        188 AlaValGluSerTyrArgLeuLeuAlaLysGlnIle--------------AspGlnPro 202
              |||::::::|||||||||||||      ::::::              |||   |||
Db        489 ATGGTCCAAGCATATCGCCTGCTTGTAGCGGAAATGTATAACCTTGGATGGGATTATCCT 548

Qy        203 LeuHisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSer 220
              ||||||||||||:::|||||||||||||          |||   :::||||||
Db        549 TTGCACTTGGGAGTTACCGAAGCTGGTGAGGGTGAAGATGGGAGGATGAAGTCT 602

RESULT 11
BJ481080
LOCUS       BJ481080                610 bp    mRNA    linear   EST 23-MAY-2002
DEFINITION  BJ481080 K. Sato unpublished cDNA library, strain H602 adult,
            heading stage top three leaves Hordeum vulgare subsp. spontaneum
            cDNA clone bah58i01 5', mRNA sequence.
ACCESSION   BJ481080
VERSION     BJ481080.1  GI:21159548
KEYWORDS    EST.
```

```
SOURCE      Hordeum vulgare subsp. spontaneum
  ORGANISM  Hordeum vulgare subsp. spontaneum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 610)
  AUTHORS   Sato,K., Saisho,D. and Takeda,K.
  TITLE     Barley EST sequencing project in NIG and Okayama Univ
  JOURNAL   Unpublished
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .610
                     /organism="Hordeum vulgare subsp. spontaneum"
                     /mol_type="mRNA"
                     /strain="H602"
                     /db_xref="taxon:77009"
                     /clone="bah58i01"
                     /tissue_type="top three leaves"
                     /dev_stage="adult, heading stage"
                     /clone_lib="K. Sato unpublished cDNA library, strain H602
                     adult, heading stage top three leaves"
BASE COUNT      176 a    118 c    155 g    160 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:              1.08e-20        Length:         610
Score:                  269.50          Matches:        68
Percent Similarity:     51.49%          Conservative:   36
Best Local Similarity:  33.66%          Mismatches:     69
Query Match:            14.44%          Indels:         29
DB:                     12              Gaps:           5

US-09-921-992-78 (1-372) x BJ481080 (1-610)

Qy         18 GlyAsnValProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArg 37
              |||||||||   :::|||        |||:::   :::|||:::|||||||   :::
Db         10 GGGAACGTGGCACTCGGCAGTGATNACCCCATGAGGATTCAGACTATGACTACCTCAGAT 69

Qy         38 ThrThrAspValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAsp 57
              |||   ||||||      ||||||   ::::::     :::          ||||||||
Db         70 ACCAAGGATGTTGCCAAGACTGTGGAAGAGGTGATGAGGATTGCAGACAGAGGCGCTGAT 129

Qy         58 IleValArgValSerValProThrMetAspAlaAlaGluAlaPhe--------------- 72
                 ||||||:::::|||               |||:::|||
Db        130 TTTGTTAGAATAACCGTCCAGGGTAAAAAGGAAGCTGATGCCTGCTTTGAGATTAAGAAC 189

Qy         73 LysLeuIleLysGlnGlnValAsnValProLeuValAlaAspIleHisPheAspTyrArg 92
                 |||:::::::::      |||:::|||||||||||||||||||||||||
Db        190 ACTCTTGTCCAGAAGAATTACAACATCCCTCTAGTGGCCGATATCCATTTTGCGCCTACA 249

Qy         93 IleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArgIleAsnProGlyAsnIle 112
              :::||||||:::|||||||||        |||   :::|||:::|||||||||||||
Db        250 GTAGCTTTAAGAGTGGCTGAATGC---TTTGACAAAATCCGTGTTAACCCAGGAAACTTT 306

Qy        113 GlyAsnGlu-------------------------------------------------- 115
                 :::
Db        307 GCCGATCGCCGTGCCCAATTTGAAAAGCTGGAATATACTGAAGACGATTACGAAAAGGAG 366

Qy        116 ---GluArgIleArgMetVal------------ValAspCysAlaArgAspLysAsnIle 130
                 |||   |||      |||             |||:::      :::
Db        367 CTTGAACACATTGAGAGGGTCTTTTCTCCATTGGTTGAGAAATGCAAGAAGTATGGAAGA 426

Qy        131 ProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGlnGluLysTyrGly 150
              :::||||||||||   |||   |||||||||        :::          ||||||
Db        427 GCCATGCGTATCGGAACAAATCATGGTAGTCTTTCTGACCGGATAATGAGCTACTATGGT 486

Qy        151 GluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeuAspArg 170
```

```
              :::   :::|||:::   ::::::|||||||||:::
Db        487 GAT---TCTCCAAGGGGAATGGTTGAGTCTGCTTTGGAATTTGCTAGGATCTGTCGGAAT 543

Qy        171 LeuAsnPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAlaValGlu 190
              |||:::|||      |||      |||:::|||||||||:::      :::   |||:::
Db        544 TTGGACTTCCATAACTTTGTATTTTCAATGAAAGCAAGTAACCCTGTTGTCATGGTCCAA 603

Qy        191 SerTyr 192
                 ::::||
Db        604 GCATAT 609

RESULT 12
BG591263
LOCUS       BG591263                706 bp    mRNA    linear   EST 07-MAR-2003
DEFINITION  EST499105 P. infestans-challenged leaf Solanum tuberosum cDNA clone
            BPLI8G5 5' sequence, mRNA sequence.
ACCESSION   BG591263
VERSION     BG591263.1  GI:13609403
KEYWORDS    EST.
SOURCE      Solanum tuberosum (potato)
  ORGANISM  Solanum tuberosum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; lamiids; Solanales; Solanaceae; Solanum.
REFERENCE   1  (bases 1 to 706)
  AUTHORS   Zhang,P., Hernandez,M., Tornqvist,C.-E, Wirtz,U., Loukoianov,A.,
            Rangel,P., Haberlach,G.T., Cho,J., Chiemingo,A., Bougri,O., Buell
            ,C.R., Ronning,C.M., Helgeson,J. and Baker,B.
  TITLE     Generation of ESTs from Potato Leaves Challenged with Phytophthora
            infestans, Incompatible Reaction
  JOURNAL   Unpublished
COMMENT     Contact: Robin Buell
            The Institute for Genomic Research
            9712 Medical Center Dr, Rockville, MD 20850, USA
            Email: potato-array@tigr.org
            This clone can be obtained from the University of Arizona Genomics
            Institute. Orders can be made through URL:
            http://genome.arizona.edu/orders/
            Seq primer: M13F-R.
FEATURES             Location/Qualifiers
     source          1. .706
                     /organism="Solanum tuberosum"
                     /mol_type="mRNA"
                     /cultivar="Kennebec"
                     /db_xref="taxon:4113"
                     /clone="BPLI8G5"
                     /tissue_type="leaf"
                     /dev_stage="6 week old"
                     /lab_host="SOLR"
                     /clone_lib="P. infestans-challenged leaf"
                     /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
                     XhoI; Whole plants were challenged with 450,000
                     sporangia/ml P. infestans US-1(US 940501) in Biotron
                     (Madison, Wisconsin). Leaf tissue was collected at 1, 2,
                     5, 12, and 24 hours post-challenge and frozen in liquid
                     nitrogen immediately upon removal. Kennebec plants showed
                     no signs of HR. Katahdin plants (susceptible to P.
                     infestans US-1) were used as controls and showed
                     infection. NOTE: We cannot exclude the possibility that
                     this sequence is actually derived from Phytophthora rather
                     than potato."
BASE COUNT      200 a    123 c    196 g    187 t
ORIGIN

Alignment Scores:
Pred. No.:              4.39e-20        Length:         706
Score:                  265.00          Matches:        65
Percent Similarity:     56.71%          Conservative:   28
Best Local Similarity:  39.63%          Mismatches:     63
Query Match:            14.20%          Indels:         8
DB:                     10              Gaps:           3

US-09-921-992-78 (1-372) x BG591263 (1-706)
```

```
Qy         90 AspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArgIleAsnPro 109
              |||   |||         :::      ||||||      |||      :::
Db         13 GACAGGCGAGCCCAGTTTGAGCAATTAGAGTACACAGAAGATGACTATCAGAAAGAACTC 72

Qy        110 GlyAsnIleGlyAsnGluGluArgIleArgMetValValAspCysAlaArgAspLysAsn 129
              :::|||      ||||||           :::|||:::      :::
Db         73 GAGCATATT------GAGGAGGTTTTTACACCATTGGTGGAAAAATGTAAGAAGTATGGA 126

Qy        130 IleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGlnGluLysTyr 149
              :::||||||||   |||   |||||||||         :::          |||
Db        127 CGTGCAATGCGCATTGGGACAAACCATGGGAGCCTTTCAGATCGCATTATGAGCTATTAT 186

Qy        150 GlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeuAsp 169
              |||:::   :::|||8::   :::::::||||||||
Db        187 GGGGAC---TCGCCTAGGGGAATGGTAGAATCAGCATTTGAGTTTGCAAGAATTTGTAGA 243

Qy        170 ArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAlaVal 189
              :::|||:::|||    |||      |||:::||||||||:::      :::   |||
Db        244 AAGTTGGACTTTCACAATTTCGTCTTCTCAATGAAAGCTAGCAATCCAGTAGTTATGGTT 303

Qy        190 GluSerTyrArgLeuLeuAlaLysGlnIle--------------AspGlnProLeuHis 204
              ::::::|||||||||||||||      ::::::              |||   |||||||||
Db        304 CAGGCGTATCGCCTTCTTGTAGCTGAGATGTATGTTCAGGGATGGGACTATCCTTTACAC 363

Qy        205 LeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeu 224
              ||||||:::|||||||||||||           |||   :::|||||||||||||||:::
Db        364 TTGGGAGTTACTGAAGCTGGTGAAGGTGAGGATGGAAGGATGAAGTCTGCAATCGGTATT 423

Qy        225 GlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspPro 244
              |||   ||||||   8:|||:::|||||||||:::||||||||||||          |||
Db        424 GGCACACTTCTTCAGGATGGTTTGGGTGATACAATACGGGTTTCCCTAACTGAAGCTCCA 483

Qy        245 ValGluGluIle 248
              |||||||||
Db        484 GAAGAGGAGATA 495

RESULT 13
AU186794
ID   AU186794   standard; RNA; EST; 489 BP.
XX
AC   AU186794;
XX
SV   AU186794.1
XX
DT   23-MAY-2003 (Rel. 75, Created)
DT   23-MAY-2003 (Rel. 75, Last updated, Version 1)
XX
DE   Porphyra yezoensis cDNA, clone:PF004b08_r, 5'end.
XX
KW   EST (expressed sequence tag).
XX
OS   Porphyra yezoensis
OC   Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyra.
XX
RN   [1]
RP   1-489
RA   Asamizu E., Nakamura Y.;
RT   ;
RL   Submitted (15-JUN-2001) to the EMBL/GenBank/DDBJ databases.
RL   Erika Asamizu, Kazusa DNA Research Institute, The First Laboratory for
RL   Plant Gene Research; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
RL   (E-mail:asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/,
RL   Tel:81-438-52-3935(ex.2336), Fax:81-438-52-3934)
XX
RN   [2]
RA   Asamizu E., Nakajima M., Kitade Y., Saga N., Nakamura Y., Tabata S.;
RT   "COMPARISON OF RNA EXPRESSION PROFILES BETWEEN THE TWO GENERATIONS OF
RT   PORPHYRA YEZOENSIS (RHODOPHYTA), BASED ON EXPRESSED SEQUENCE TAG FREQUENCY
RT   ANALYSIS";
RL   J. Phycol. 0:0-0(2003).
XX
FH   Key             Location/Qualifiers
FH
FT   source          1. .489
FT                   /db_xref="taxon:2788"
FT                   /mol_type="mRNA"
FT                   /organism="Porphyra yezoensis"
FT                   /clone="PF004b08_r"
FT                   /dev_stage="sporophytes"
FT                   /strain="TU-1"
XX
SQ   Sequence 489 BP; 100 A; 136 C; 170 G; 83 T; 0 other;

Alignment Scores:
Pred. No.:              3.42e-20        Length:         489
Score:                  264.00          Matches:        68
Percent Similarity:     57.14%          Conservative:   24
Best Local Similarity:  42.24%          Mismatches:     48
Query Match:            14.15%          Indels:         22
DB:                     6               Gaps:           5

US-09-921-992-78 (1-372) x AU186794 (1-489)

Qy        104 CysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMetValValAsp 123
              |||||||||   :::|||         |||         |||   |||
Db         14 TGTTTGAGA---AGCCCAAGGAGGGCCGGACGGAGTACTCGGACGAGGAGTTCAACCGGC 70

Qy        124 ----------------------------------CysAlaArgAspLysAsn------ 129
                                                   |||   |||   |||:::
Db         71 TCAAAGACAAGCTCAAGGCCACCCTCACACCACTGGTGTTGTCGCTCAAAGAGCAGGGCA 130

Qy        130 IleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGlnGluLysTyr 149
              |||   ||||||||||||||   |||||||||   :::   :::         |||
Db        131 AGGCCATG-CGTATTGGGGTCAACCACGGGTCCCTGGCCGAGCGCATGATGTTCACCTAC 189

Qy        150 GlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeuAsp 169
              |||:::   |||||||:::   ::::::||||||||||||      ::::::
Db        190 GGGGAC---ACCCCGCGGGGGATGGTCGAGTCGGCGATGGAGTGCATTGAGATCTGCCGG 246

Qy        170 ArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAlaVal 189
              |||:::|||            ::::|||:::|||||||||   |||   :::   :::
Db        247 GAGCTAGACTTTCACAACCTGGTCATTTCGATGAAGGCGTCCAAGGTGCCAGTCATGATC 306

Qy        190 GluSerTyrArgLeuLeuAlaLysGlnIle--------------AspGlnProLeuHis 204
              :::|||||   ||||||   ::::::              |||   |||:::|||
Db        307 GCAGCCTACCGCCAGCTGGCGCTGCGGCTCGACGCTGAGGGGTACGACTACCCCATCCAC 366

Qy        205 LeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeu 224
              ||||||:::|||||||||||||           |||   |||||||||:::||||||:::
Db        367 CTGGGTGTGACTGAGGCGGGCGACGGGGAGTATGGGCGGGTGAAGAGCTCCATCGGTATC 426

Qy        225 GlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspPro 244
              |||   ||||||   ||||||||||||||||||||||||||||||||      :::|||
Db        427 GGCACGCTCCTTGGGGAGGGCATCGGCGACACGCTGCGGGTGAGCTTGACAGAGAACCCC 486

Qy        245 Val 245
              :::
Db        487 ATG 489

RESULT 14
AW042702
LOCUS       AW042702                601 bp    mRNA    linear   EST 18-SEP-1999
DEFINITION  ST23E04 Pine TriplEx shoot tip library Pinus taeda cDNA clone
            ST23E04, mRNA sequence.
ACCESSION   AW042702
VERSION     AW042702.1  GI:5903147
KEYWORDS    EST.
SOURCE      Pinus taeda (loblolly pine)
  ORGANISM  Pinus taeda
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.
REFERENCE   1  (bases 1 to 601)
  AUTHORS   Whetten,R.W., Kinlaw,C.S., Retzel,E. and Sederoff,R.R.
  TITLE     The Pine Gene Discovery Project
  JOURNAL   Unpublished
```

```
COMMENT      Contact: Ross Whetten
             Forest Biotechnology Group
             North Carolina State University
             Dept. of Forestry, NC State University, 6113 Jordan Hall, Raleigh
             ,NC, 27695-8008
             Tel: 919-515-7800
             Fax: 919-515-7801
             Email: rosswhet@unity.ncsu.edu
             Seq primer: 5' lambda TriplEx2 Sequencing Primer.
FEATURES              Location/Qualifiers
     source           1. .601
                      /organism="Pinus taeda"
                      /mol_type="mRNA"
                      /db_xref="taxon:3352"
                      /clone="ST23E04"
                      /lab_host="E. coli BM25.8"
                      /clone_lib="Pine TriplEx shoot tip library"
                      /note="Organ: shoot tips; Vector: Lambda TriplEx; Site_1:
                      SfiI (A); Site_2: SfiI (B); Shoot tips (approx. 2 cm from
                      apex) were collected during the spring, frozen and used
                      for mRNA isolation. The SMART-PCR method (Clontech) was
                      used to prepare a library from 1 ug total RNA, using the
                      Lambda TriplEx vector. Plasmid subclones in pTriplEx were
                      recovered by cre-lox excision in E. coli strain BM25.8 and
                      sequenced from the 5' end."
BASE COUNT     162 a    103 c    166 g    165 t      5 others
ORIGIN

Alignment Scores:
Pred. No.:               7.76e-20      Length:          601
Score:                   262.00        Matches:         57
Percent Similarity:      62.30%        Conservative:    19
Best Local Similarity:   46.72%        Mismatches:      40
Query Match:             14.04%        Indels:          6
DB:                      9             Gaps:            2

US-09-921-992-78 (1-372) x AW042702 (1-601)

Qy       132 IleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGlnGluLysTyrGlyGlu 151
             :::|||||||||   |||   |||||||||                     |||||||:::
Db        24 ATGCGAATTGGCACAAACCATGGAAGTCTTTCCGATCGTACTATGAGTTATTATGGTGAT 83

Qy       152 ProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeuAspArgLeu 171
                :::|||:::   ::::::|||||||||                          :::|||
Db        84 ---TCTCCCAGGGGTATGGTGGAATCAGCATTTGAATTTGCACGCATTTGCCGGAAGTTG 140

Qy       172 AsnPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAlaValGluSer 191
                |||      |||      |||::::|||||||||||       :::    |||::::::
Db       141 GGTTTTCATAATTTTGTGTTTTCAATGAAAGCGAGCGATCCTGTAGTCATGGTTCAGGCA 200

Qy       192 TyrArgLeuLeuAlaLysGlnIle---------------AspGlnProLeuHisLeuGly 206
             ||||||||||||      ::::::               |||   |||||||||||||||
Db       201 TACCGTTTACTTGTTGCGGAGATGTATGTGCAAGGATGGGATTATCCATTGCATTTAGGA 260

Qy       207 IleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeu 226
             ·:::|||||||||||         |||   :::||||||||||||||||:::|||
Db       261 GTTACTGAAGCTGGTGAAGGTGAAGATGGACGCATGAAGTCTGCAATTGGCATTGGAACA 320

Qy       227 LeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGlu 246
             ||||||   :::|||:::|||||||||:::||||||||||||          |||   |||
Db       321 CTTTTGCAGGATGGTTTGGGTGATACTATTCGAGTTTCCCTTACAGAACCTCCAGAAGAG 380

Qy       247 GluIle 248
             ||||||
Db       381 GAGATC 386

RESULT 15
BJ551332
LOCUS       BJ551332                 603 bp    mRNA    linear   EST 14-NOV-2002
DEFINITION  BJ551332 K. Sato unpublished cDNA library, strain H602 adult,
            heading stage top three leaves Hordeum vulgare subsp. spontaneum
            cDNA clone bah59b20 5', mRNA sequence.
ACCESSION   BJ551332
VERSION     BJ551332.1  GI:24969783
KEYWORDS    EST.
SOURCE      Hordeum vulgare subsp. spontaneum
  ORGANISM  Hordeum vulgare subsp. spontaneum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 603)
  AUTHORS   Sato,K., Saisho,D. and Takeda,K.
  TITLE     Barley EST sequencing project in NIG and Okayama Univ
  JOURNAL   Unpublished
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .603
                     /organism="Hordeum vulgare subsp. spontaneum"
                     /mol_type="mRNA"
                     /strain="H602"
                     /db_xref="taxon:77009"
                     /clone="bah59b20"
                     /tissue_type="top three leaves"
                     /dev_stage="adult, heading stage"
                     /clone_lib="K. Sato unpublished cDNA library, strain H602
                     adult, heading stage top three leaves"
BASE COUNT     156 a    112 c    169 g    165 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:               1.02e-19      Length:          603
Score:                   261.00        Matches:         56
Percent Similarity:      63.93%        Conservative:    22
Best Local Similarity:   45.90%        Mismatches:      38
Query Match:             13.99%        Indels:          6
DB:                      12            Gaps:            2

US-09-921-992-78 (1-372) x BJ551332 (1-603)

Qy       132 IleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGlnGluLysTyrGlyGlu 151
             :::|||||||||   |||   |||||||||         :::          ||||||:::
Db        42 CTGCGTATCGGAACAAATCATGGTAGTCTTTCTGACCGGATAATGAGCTACTATGGTGAT 101

Qy       152 ProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeuAspArgLeu 171
                :::|||:::   ::::::|||||||||:::                           |||
Db       102 ---TCTCCAAGGGGAATGGTTGAGTCTGCTTTGGAATTTGCTAGGATCTGTCGGAATTTG 158

Qy       172 AsnPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAlaValGluSer 191
             :::|||      |||      |||:::|||||||||:::       :::   |||::::::
Db       159 GACTTCCATAACTTTGTATTTTCAATGAAAGCAAGTAACCCTGTTGTCATGGTCCAAGCA 218

Qy       192 TyrArgLeuLeuAlaLysGlnIle---------------AspGlnProLeuHisLeuGly 206
             ||||||||||||      ::::::               |||   |||||||||||||||
Db       219 TATCGCCTGCTTGTAGCGGAAATGTATAACCTTGGATGGGATTATCCTTTGCACTTGGGA 278

Qy       207 IleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeu 226
             :::||||||||||||         |||   :::||||||||||||||||:::|||
Db       279 GTTACCGAAGCTGGTGAGGGTGAAGATGGGAGGATGAAGTCTGCTATTGGCATCGGAACA 338

Qy       227 LeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGlu 246
             ||||||   :::|||:::|||||||||:::||||||||||||          |||   |||
Db       339 CTTTTGATGGATGGCTTGGGTGATACAATCCGTGTATCCCTCACAGAACCACCAGAGGAA 398

Qy       247 GluIle 248
             ||||||
Db       399 GAAATT 404

Search completed: August 15, 2003, 01:35:09
Job time : 2639 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 12, 2003, 09:54:00 ; Search time 98 Seconds
(without alignments)
979.547 Million cell updates/sec

Title: US-09-921-992-78
Perfect score: 1866
Sequence: 1 MHNQAPIQRRKSTRIYVGNV.....RAKASQLDEARRIDVQQVEK 372

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_23:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1631.5	87.4	372	16	Q8DEZ8	Q8dez8 vibrio vuln
2	1584.5	84.9	371	16	Q8EC32	Q8ec32 shewanella
3	1226.5	65.7	366	16	Q8D1Y3	Q8d1y3 wiggleswort
4	814.5	43.6	403	16	Q8G7Y6	Q8g7y6 bifidobacte
5	753	40.4	393	16	Q8FP82	Q8fp82 corynebacte
6	739.5	39.6	362	16	Q8EUI6	Q8eui6 mycoplasma
7	601.5	32.2	402	16	Q8DK70	Q8dk70 synechococc
8	525.5	28.2	663	16	Q8F1H5	Q8f1h5 leptospira
9	508	27.2	420	16	Q8FYT2	Q8fyt2 brucella su
10	471	25.2	740	10	Q8GZR6	Q8gzr6 lycopersico
11	463	24.8	741	10	Q8RXG8	Q8rxg8 arabidopsis
12	462	24.8	716	10	Q9FF59	Q9ff59 arabidopsis
13	462	24.8	740	10	Q8GZR7	Q8gzr7 arabidopsis
14	457	24.5	741	10	Q8LPQ4	Q8lpq4 arabidopsis
15	455.5	24.4	746	16	Q8KG23	Q8kg23 chlorobium
16	446	23.9	824	5	Q9BJX5	Q9bjx5 plasmodium
17	446	23.9	824	5	Q8IJH7	Q8ijh7 plasmodium
18	136	7.3	803	16	Q8R927	Q8r927 thermoanaer
19	122.5	6.6	507	17	Q8TZP6	Q8tzp6 pyrococcus
20	119	6.4	535	16	Q8DJ32	Q8dj32 synechococc
21	114.5	6.1	588	2	Q8VSQ9	Q8vsq9 actinobacil
22	114	6.1	524	17	O27773	O27773 methanobact
23	112.5	6.0	686	17	Q9UXG1	Q9uxg1 sulfolobus
24	112.5	6.0	723	3	Q9P4D5	Q9p4d5 emericella
25	111.5	6.0	344	16	Q8EQA9	Q8eqa9 oceanobacil
26	111.5	6.0	578	2	O68221	O68221 anaplasma p
27	111.5	6.0	588	2	Q8VSR0	Q8vsr0 actinobacil
28	111	5.9	575	16	Q8XVK5	Q8xvk5 ralstonia s
29	109	5.8	764	2	Q9RNG5	Q9rng5 legionella
30	108.5	5.8	449	16	Q8PJZ9	Q8pjz9 xanthomonas
31	108.5	5.8	1000	17	Q8TPA1	Q8tpa1 methanosarc
32	106.5	5.7	284	16	Q9CHV6	Q9chv6 lactococcus
33	106.5	5.7	798	2	Q9KHS7	Q9khs7 pseudomonas
34	106.5	5.7	798	2	Q8KUR5	Q8kur5 pseudomonas
35	106	5.7	748	17	Q57616	Q57616 methanococc
36	105.5	5.7	578	2	Q93NY9	Q93ny9 anaplasma p
37	104.5	5.6	638	5	Q9BMQ6	Q9bmq6 opisthorchi
38	104	5.6	606	16	Q8Y7W3	Q8y7w3 listeria mo
39	104	5.6	1227	16	Q97K41	Q97k41 clostridium
40	103.5	5.5	537	16	Q8FL75	Q8fl75 escherichia
41	102.5	5.5	417	3	Q9C473	Q9c473 emericella
42	102	5.5	546	16	Q8DZ63	Q8dz63 streptococc
43	102	5.5	546	16	Q8DX35	Q8dx35 streptococc
44	101.5	5.4	564	5	O44422	O44422 heliothis v
45	101.5	5.4	572	10	Q8GZY8	Q8gzy8 oryza sativ

ALIGNMENTS

RESULT 1
Q8DEZ8
ID Q8DEZ8 PRELIMINARY; PRT; 372 AA.
AC Q8DEZ8;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE GcpE.
GN VV10427.
OS Vibrio vulnificus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=672;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CMCP6;
RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
RA Choy H.E.;
RT "Complete genome sequence of Vibrio vulnificus CMCP6.";
RL Submitted (DEC-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AE016798; AAO08950.1; -.
KW Complete proteome.
SQ SEQUENCE 372 AA; 40573 MW; 6C73C2B8819B0285 CRC64;

Query Match 87.4%; Score 1631.5; DB 16; Length 372;
Best Local Similarity 86.7%; Pred. No. 1.3e-112;
Matches 320; Conservative 29; Mismatches 19; Indels 1; Gaps 1;

```
Qy          1 MHNQAPIQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVR 60
              |||::|| |||||||||:|||||||||||||||||| ||| |||||| |||||||
Db          1 MHNESPIIRRKSTRIYVGDVPIGDGAPIAVQSMTNTRTTDVAATVAQIKALENVGADIVR 60

Qy         61 VSVPTMDAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRM 120
              ||||||||||||||||||::|||||||||||||||||||||||||||||||||||||||
Db         61 VSVPTMDAAEAFKLIKQQVSIPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRS 120

Qy        121 VVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSV 180
              |||||||||||||||| |||||||| |||||||:||:|||||||| ||||||||||||
Db        121 VVDCARDKNIPIRIGVNGGSLEKDLQMKYGEPTPEALVESAMRHVDILDRLNFDQFKVSV 180
```

```
Qy        181 KASDVFLAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSL 240
              |||||||||:|||||||||||||||||||||||||:||||||:|||:||:|||||||:||
Db        181 KASDVFLAVDSYRLLAKKIDQPLHLGITEAGGARAGAVKSAVGLGMLLAEGIGDTLRISL 240

Qy        241 AADPVEEIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDV 300
              ||:|||||||||||||||||||||||||:|||||||| ||||||:||||||||||
Db        241 AANPVEEIKVGFDILKSLRIRSRGINFIACPSCSRQEFDVISTVNALEERLEDIITPMDV 300

Qy        301 SIIGCVVNGPGEALVSTLGVTGGNKKSGLYEDGVR-KDRLDNNDMIDQLEARIRAKASQL 359
              |||||||||||| || ||: | ||||  |||| | |:| || |:::||||:|||||:::
Db        301 SIIGCVVNGPGEAEVSHLGLAGSNKKSAFYEDGKRQKERFDNEDLVNQLEAKIRAKAARM 360

Qy        360 DEARRIDVQ 368
              ||: |||::
Db        361 DESNRIDIK 369
```

RESULT 2
Q8EC32
ID Q8EC32 PRELIMINARY; PRT; 371 AA.
AC Q8EC32;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase.
GN ISPG OR SO3312.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
OX NCBI_TaxID=70863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MR-1;
RX MEDLINE=22297686; PubMed=12368813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A.,
RA Feldblyum T.V., Smith H.O., Venter J.C., Nealson K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis.";
RL Nat. Biotechnol. 20:1118-1123(2002).
DR EMBL; AE015769; AAN56310.1; -.
DR TIGR; SO3312; -.
KW Complete proteome.
SQ SEQUENCE 371 AA; 40603 MW; 024A994C6EC93906 CRC64;

Query Match 84.9%; Score 1584.5; DB 16; Length 371;
Best Local Similarity 84.8%; Pred. No. 3.8e-109;
Matches 313; Conservative 32; Mismatches 21; Indels 3; Gaps 2;

```
Qy          1 MHNQAPIQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVR 60
              |:|: ||:|| |||||||||||||||||||||||:|||||||| ||:|||:|||||||
Db          1 MYNETPIKRRPSTRIYVGNVPIGDGAPIAVQSMTNTKTTDVEATVAQIRALEKVGADIVR 60

Qy         61 VSVPTMDAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRM 120
              |||||||||||||||| |:|||||||||||||||||||||||||||||||||||||||||
Db         61 VSVPTMDAAEAFKLIKQSVSVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRS 120

Qy        121 VVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSV 180
              ||:|||||||||||||| ||||||| :|| ||||:|||||||||| ||||||||||||
Db        121 VVECARDKNIPIRIGVNGGSLEKDLMDKYKEPTPEALLESAMRHVDILDRLNFDQFKVSV 180

Qy        181 KASDVFLAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSL 240
              |||||||||||||||||| |||||||||||||||:||||||:|||:||:|||||||:||
Db        181 KASDVFLAVESYRLLAKQIRQPLHLGITEAGGARAGAVKSAVGLGMLLAEGIGDTLRISL 240

Qy        241 AADPVEEIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDV 300
              |||||||||||||||||||||||||||||:||||||||| ||| ||:|||||: | |||
```

```
Db        241 AADPVEEIKVGFDILKSLRIRSRGINFIACPSCSRQEFDVISTVNELERRLEDVTTAMDV 300

Qy        301 SIIGCVVNGPGEALVSTLGVTGGNKKSGLYEDGVR-KDRLDNNDMIDQLEARIRAKASQL 359
              |||||||||||||||| :|:|||::||| |::| | |:| ||::::| |||:|||||||:
Db        301 SIIGCVVNGPGEALVSHIGLTGGHRKSGYYDEGERQKERFDNDNLVDSLEAKIRAKASQM 360

Qy        360 DEARRIDVQ 368
                | || |:
Db        361 --ANRIQVK 367
```

RESULT 3
Q8D1Y3
ID Q8D1Y3 PRELIMINARY; PRT; 366 AA.
AC Q8D1Y3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE GcpE protein.
GN GCPE.
OS Wigglesworthia brevipalpis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Wigglesworthia.
OX NCBI_TaxID=164609;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22297718; PubMed=12219091;
RA Akman L., Yamashita A., Watanabe H., Oshima K., Shiba T., Hattori M.,
RA Aksoy S.;
RT "Genome sequence of the endocellular obligate symbiont of tsetse
RT flies, Wigglesworthia glossinidia.";
RL Nat. Genet. 32:402-407(2002).
DR EMBL; AB063522; BAC24719.1; -.
KW Complete proteome.
SQ SEQUENCE 366 AA; 40407 MW; 159E0169E8FC8DED CRC64;

Query Match 65.7%; Score 1226.5; DB 16; Length 366;
Best Local Similarity 65.9%; Pred. No. 1.2e-82;
Matches 234; Conservative 62; Mismatches 58; Indels 1; Gaps 1;

```
Qy          7 IQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVRVSVPTM 66
              | |||| :||: |||||  :||:|||||| |||: :|::||  |:: |||||||:||:
Db          7 IIRRKSKKIYIKNVPIGGDSPISVQSMTNTCTTDINSTISQINKLQKAGADIVRVSIPTL 66

Qy         67 DAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRMVVDCAR 126
              :|||:||:||: |::|:|||||||||||||| |||| ||||||||||   || || |:
Db         67 EAAESFKIIKRNVSIPIVADIHFDYRIALKAAEYGADCLRINPGNIGKLNRIISVVSTAK 126

Qy        127 DKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSVKASDVF 186
              :| :|||||||:||||||:: |||   |::| |||||||: |::|||| ||||||:|||
Db        127 EKKLPIRIGVNSGSLEKDIENKYGINNPKSLFESAMRHVNILEKLNFDMFKVSVKSSDVL 186

Qy        187 LAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSLAADPVE 246
                |:||:||| :|||||||||||:|    |::||:||:|||||||||||||||||:|
Db        187 TCVQSYKLLASKIDQPLHLGITESGSMLHGSIKSSIGIGLLLSEGIGDTLRVSLAADPIE 246

Qy        247 EIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDVSIIGCV 306
              |:|||| ||:|| || |||||||||||||||  || ||:||||:||||:||:|||:
Db        247 EVKVGFSILRSLNIRKRGINFIACPTCSRQEFDVINVVNVLEKRLEDVITPMNVSVIGCM 306

Qy        307 VNGPGEALVSTLGVTGGNKKSGLYEDGVR-KDRLDNNDMIDQLEARIRAKASQLD 360
              ||| |||  : :|::|   || |:|:|:|  ::::| ::||:||  || |   |:
Db        307 VNGLGEANRADIGISGSRNKSILFENGLRNNNKINNEEIIDKLEKYIRKKVKILN 361
```

RESULT 4
Q8G7Y6
ID Q8G7Y6 PRELIMINARY; PRT; 403 AA.
AC Q8G7Y6;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase 1.

```
GN   ISPG OR BL0098.
OS   Bifidobacterium longum.
OC   Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
OC   Bifidobacteriaceae; Bifidobacterium.
OX   NCBI_TaxID=216816;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCC 2705;
RX   MEDLINE=22294977; PubMed=12381787;
RA   Schell M.A., Karmirantzou M., Snel B., Vilanova D., Berger B.,
RA   Pessi G., Zwahlen M.-C., Desiere F., Bork P., Delley M.,
RA   Pridmore R.D., Arigoni F.;
RT   "The genome sequence of Bifidobacterium longum reflects its adaptation
RT   to the human gastrointestinal tract.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427(2002).
DR   EMBL; AE014624; AAN23963.1; -.
KW   Complete proteome.
SQ   SEQUENCE   403 AA;  42564 MW;  F6F8C3A6153A6056 CRC64;

  Query Match             43.6%;  Score 814.5;  DB 16;  Length 403;
  Best Local Similarity   46.3%;  Pred. No. 4.2e-52;
  Matches  167;  Conservative   70;  Mismatches  111;  Indels   13;  Gaps    4;

Qy          3 NQAPIQ-RRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVRV 61
              :::|:  |||| || || ||:| ||||:|||||| | :| ||: ||  |   | |||||
Db         22 SESPLHPRRKSRRIMVGPVPVGGGAPISVQSMTNTLTANVPATLQQIAELTAAGCDIVRV 81

Qy         62 SVPTMDAAEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRM- 120
              :||: | |:|   | ::  :|::|||||  :   :  : |   :|:|||||   : :
Db         82 AVPSQDDADALPEICRKSPIPVIADIHFQSKYVFQAIDAGCAAVRVNPGNIRKFDEVGPD 141

Qy        121 VVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSV 180
              :   | |  | :|||||||||:|:|  ||| |||:||: ||::     : : |  ||:||
Db        142 ICKAATDAGISLRIGVNAGSLDKELYAKYGGPTPEALVASALKEAHMFEDVGFHDFKISV 201

Qy        181 KASDVFLAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSL 240
              |  ||   ||:|||| : | |||||:|||| |  | :|| :  | ||:|||||:|||
Db        202 KHHDVITMVETYRLLASKGDWPLHLGVTEAGPAWQGTIKSCLAFGALLAEGIGDTIRVSL 261

Qy        241 AADPVEEIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDV 300
              :| | ||:|||  :|: : :| |  : |:||:| | : |||   :|: : |:|:  |: |
Db        262 SAPPAEEVKVGCKLLEYMGLRPRKFDIISCPSCGRAQVDVIQLASAVTEGLKDVTAPIRV 321

Qy        301 SIIGCVVNGPGEALVSTLGVTGGNKKSGLYEDG-----VRKDRLDN------NDMIDQLE 349
              :::||:|||||||  : |||  || |  ::  |     | :|::  :      ||:  |:|
Db        322 AVMGCIVNGPGEAREADLGVASGNGKGQIFIKGKVIKTVPEDQIVDTLLTIANDIAAQME 381

Qy        350 A 350
              |
Db        382 A 382


RESULT 5
Q8FP82
ID   Q8FP82      PRELIMINARY;      PRT;   393 AA.
AC   Q8FP82;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Putative aminoglycoside acetyltransferase negative regulator.
GN   CE1903.
OS   Corynebacterium efficiens.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX   NCBI_TaxID=152794;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;
RA   Kawarabayasi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,
RA   Ikeo K., Suzuki M., Mashima J., Itoh T., Yamagishi A., Nishio Y.,
RA   Usuda Y., Sugimoto S.;
RT   "The entire genomic sequence of Corynebacterium efficiens YS-314.";
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP005220; BAC18713.1; -.
KW   Transferase; Complete proteome.
SQ   SEQUENCE   393 AA;  41533 MW;  A8CC2A98F150AD66 CRC64;

  Query Match             40.4%;  Score 753;  DB 16;  Length 393;
  Best Local Similarity   43.7%;  Pred. No. 1.4e-47;
  Matches  157;  Conservative   67;  Mismatches  123;  Indels   12;  Gaps    3;

Qy          9 RRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVRVSVPTMDA 68
              |||: :: ||:| :|   ||:||||| |:| |: ||: ||  |   | |||||: |
Db         20 RRKTRQLMVGSVGVGSDHPISVQSMTTTKTHDINATLQQIAQLTASGCDIVRVACPKPVD 79

Qy         69 AEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGN-EERIRMVVDCARD 127
              ||| :| ::  :|::|||||  :      : |   :|:|||||   : |:: |   | |
Db         80 AEALPIIAKKSPIPVIADIHFQPKYIFSAIDAGCAAVRVNPGNIKEFDGRVKEVAQAAGD 139

Qy        128 KNIPIRIGVNAGSLEKDLQEKY-GEPTPQALLESAMRHVDHLDRLNFDQFKVSVKASDVF 186
              |||||||| |||:| : :|| |: ||:||:|||:      :   :    :||| ||
Db        140 AGIPIRIGVNGGSLDKRILDKYHGKATPEALVESALWEASLFEEHGYGDIAISVKHSDPV 199

Qy        187 LAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSLAADPVE 246
              | ||:|| ||:: | |||||:||||    | :||::  | |||:|||||:|||||:|||||
Db        200 LMVEAYRQLAEKCDYPLHLGVTEAGPKFMGTIKSSVAFGALLSQGIGDTIRVSLSADPVE 259

Qy        247 EIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDVSIIGCV 306
              |||||  ||:|| :| | :  ::||:| | : ||      : : |: :  |: |:::|||
Db        260 EIKVGDQILQSLNLRPRKLEIVSCPSCGRAQVDVYKLAEEVTEGLDGLEVPLRVAVMGCV 319

Qy        307 VNGPGEALVSTLGVTGGNKKSGLYEDGVRKDRLDNNDMIDQLEARIRAKASQLDEARRI 365
              |||||||  : |||  || |  ::  |     : : ::  |          ::|| ||
Db        320 VNGPGEARDADLGVASGNGKGQIFVKGEIIKTVPESQIVQTL----------IEEAMRI 368


RESULT 6
Q8EUI6
ID   Q8EUI6      PRELIMINARY;      PRT;   362 AA.
AC   Q8EUI6;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Peptidoglycan acetylation.
GN   MYPE9400.
OS   Mycoplasma penetrans.
OC   Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX   NCBI_TaxID=28227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HF-2;
RX   MEDLINE=22354719; PubMed=12466555;
RA   Sasaki Y., Ishikawa J., Yamashita A., Oshima K., Kenri T., Furuya K.,
RA   Yoshino C., Horino A., Shiba T., Sasaki T., Hattori M.;
RT   "The complete genomic sequence of Mycoplasma penetrans, an
RT   intracellular bacterial pathogen in humans.";
RL   Nucleic Acids Res. 30:5293-5300(2002).
DR   EMBL; AP004174; BAC44727.1; -.
KW   Complete proteome.
SQ   SEQUENCE   362 AA;  39237 MW;  3F7EB12EF331AE73 CRC64;

  Query Match             39.6%;  Score 739.5;  DB 16;  Length 362;
  Best Local Similarity   41.0%;  Pred. No. 1.3e-46;
  Matches  146;  Conservative   82;  Mismatches  127;  Indels    1;  Gaps    1;

Qy          9 RRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVRVSVPTMDA 68
              | |:  :|||| ||    : :|||| |:| ||| || |:| | | | ::||::|   :
Db          6 REKTKAVYVGNVQIGGNNKVVIQSMTTTKTHDVEKTVAQVKELVREGCELVRIAVLDDED 65

Qy         69 AEAFKLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRMVVDCARDK 128
              | ||  : :    |::|||||:   |:|  | |   :|:|||| :||::| ::| |  |
Db         66 AAAFGEVVKNSPCPIIADIHFNPLYAIKAIESGAAKVRLNPGNIKDEEQLRKIIDLANKK 125

Qy        129 NIPIRIGVNAGSLEKDLQEKYGEPTPQALLESAMRHVDHLDRLNFDQFKVSVKASDVFLA 188
              |||||:|||:|||  || : :|  |  |:: :  |::  :   |:  ||:||::| ||
Db        126 NIPIRVGVNSGSLPMDLMKSHG-VTADAMMIAVKRYINLFESNGFNNIVVSLKATNVLLA 184
```

```
Qy        189 VESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSLAADPVEEI 248
              :|:|: :| : : |||:|||||   :| :||| |||:|| ||||:|:|:||  ||: |:
Db        185 IEAYKKAAMEFNYPLHIGITEAGSLFNGTIKSAAGLGVLLHEGIGNTIRISLTGDPLSEV 244

Qy        249 KVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDVSIIGCVVN 308
              ||   :| || :    :: |:|||| |  ||:   |  :|:    :  |: |:|:|| ||
Db        245 KVCKKLLNSLGLYDNLVDIISCPTCGRLNFDLNPVVKEIEKFTRKMNFPLKVAILGCAVN 304

Qy        309 GPGEALVSTLGVTGGNKKSGLYEDGVRKDRLDNNDMIDQLEARIRAKASQLDEARR 364
              |||||  : :|: |||    :: :|    :  : ::|:|:  |  |  :  ::::
Db        305 GPGEAKEADIGIAGGNGTGIIFANGKAIKSVPEDQLVDELKKLISIKYKEYLDSKK 360


RESULT 7
Q8DK70
ID   Q8DK70      PRELIMINARY;      PRT;   402 AA.
AC   Q8DK70;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Tlr0996 protein.
GN   TLR0996.
OS   Synechococcus elongatus (Thermosynechococcus elongatus).
OC   Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX   NCBI_TaxID=32046;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BP-1;
RX   MEDLINE=22225144; PubMed=12240834;
RA   Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
RA   Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,
RA   Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
RA   Shimpo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
RT   "Complete genome structure of the thermophilic cyanobacterium
RT   Thermosynechococcus elongatus BP-1.";
RL   DNA Res. 9:123-130(2002).
DR   EMBL; AP005372; BAC08548.1; -.
KW   Complete proteome.
SQ   SEQUENCE   402 AA;  44192 MW;  7595FFC940AC60D7 CRC64;

  Query Match             32.2%;  Score 601.5;  DB 16;  Length 402;
  Best Local Similarity   36.8%;  Pred. No. 2.4e-36;
  Matches  147;  Conservative   79;  Mismatches  120;  Indels   53;  Gaps   11;

Qy          4 QAPIQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVRVSV 63
              :  | |||:  : :|30 || | |:|||| |  | |:| :|  |: |  :| :||||:|
Db         14 ETAIVRRKTRPVPIGSVVIGGGHPVAVQSMINEDTLDIEGSVAAIRRLHEIGCEIVRVTV 73

Qy         64 PTMDAAEAF-----KLIKQQVNVPLVADIHFD-YRIALKVAEYGVDCLRINPG------- 110
              |::  |:|      :|||    ||||||:| :  :|||:||:| || :|||||
Db         74 PSLAHAKAMEEIRDRLYKTYKPVPLVADVHHNGMKIALEVAKY-VDNVRINPGLYVFEKP 132

Qy        111 -------------NIGNE--ERIRMVVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQ 155
                          |||:  | :  :|   ||:   :|||||| ||| : :   ||: ||:
Db        133 KPNRTEYTQAEFDEIGAKIRETLEPLVISLRDQGKSMRIGVNHGSLAERMLFTYGD-TPE 191

Qy        156 ALLESAMRHVDHLDRLNFDQFKVSVKASDVFLAVESYRLLAKQIDQ-----PLHLGITEA 210
              ::|||:  :   : 10|   ::|:||| | : : : ||: |::|:     |||||:|||
Db        192 GMVESALEFIRICESLNFYNLEISLKASRVPVMIAANRLMVKRMDELGMDYPLHLGVTEA 251

Qy        211 GGARSGAVKSAIGLGLLLSEGIGDTLRVSLAADPVEEIKVGFDILKSLRIRSRGINFIAC 270
              |    | :||  |: 0|:||||||:||||   | :|| | : ||::| :|   : ::||
Db        252 GDGEYGRIKSTAGIALLAEGIGDTIRVSLTEAPEKEIPVCYGILQALGLRRTMVEYVAC 311

Qy        271 PTCSRQEFDVIGTVNALEQRLEDI------ITPMDVSIIGCVVNGPGEALVSTLGVTGGN 324
              |:| |  |:      0|: |  :      :| ::::::||:||||||   :  |  |
Db        312 PSCGRTLFN------LEEVLHKVREATKHLTGLNIAVMGCIVNGPGEMADADYGYVG-- 362

Qy        325 KKSG---LYEDGVRKDRLDNNDMIDQLEARIRAKASQLD 360
              |: |   ||       8:   : :  |   |:|    :|
Db        363 KQPGYISLYRGREEVKKVPEAEGVAALVELIKADGRWVD 401
```

```
RESULT 8
Q8F1H5
ID   Q8F1H5      PRELIMINARY;      PRT;   663 AA.
AC   Q8F1H5;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   GcpE protein homolog.
GN   GCPE OR LA3160.
OS   Leptospira interrogans.
OC   Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.
OX   NCBI_TaxID=173;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=56601 / Serogroup Icterohaemorrhagiae / Serovar lai;
RA   Ren S.;
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE011478; AAN50358.1; -.
KW   Complete proteome.
SQ   SEQUENCE   663 AA;  73779 MW;  79B600DF805A6365 CRC64;

  Query Match             28.2%;  Score 525.5;  DB 16;  Length 663;
  Best Local Similarity   41.7%;  Pred. No. 2.1e-30;
  Matches  118;  Conservative   53;  Mismatches   75;  Indels   37;  Gaps    8;

Qy          2 HNQAPI--QRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIV 59
              :|| |   ||||:  : ||:| :|   || :||| |: ||| : :| ||  ||| | :||
Db          5 YNQTPFGYQRRKTREVKVGDVKVGGNNPIVIQSMINSDTTDTQGSVKQILELERAGCEIV 64

Qy         60 RVSVPTMDAAEAFKLIKQQV-----NVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGN 114
              |:::||:   |:    |:|::      |||||||||   :|:|  || |: :||||||  :
Db         65 RLTVPSQADADNLPSIRQELKKAGSKVPLVADIHFTPSVAMKAVEY-VEKVRINPGNFAD 123

Qy        115 E-----------------ERIR-----MVVDCARDKNIPIRIGVNAGSLEKDLQEKYGEP 152
              :                 |||      :|: | ::  : :||| | |||   :  :||:
Db        124 KKKFAVRDYTDLEYNQELERISEVFSPLVLRC-KELGVSMRIGTNHGSLSDRIMNRYGD- 181

Qy        153 TPQALLESAMRHVDHLDRLNFDQFKVSVKASDVFLAVESYRLLAK-----QIDQPLHLGI 207
              ||| ::|||:  :   : | :    ||:|||:  : |::||:||      ::| |||||:
Db        182 TPQGMVESALEFIRIAESLGYYDIIVSMKASNPQVMVQAYRMLASRFNELKMDYPLHLGV 241

Qy        208 TEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSLAADPVEEIKV 250
              ||||    | :|||||:| || :|:|||:||||  ||| |: |
Db        242 TEAGDGNDGRIKSAIGIGSLLEDGLGDTIRVSLTEDPVLEVPV 284


RESULT 9
Q8FYT2
ID   Q8FYT2      PRELIMINARY;      PRT;   420 AA.
AC   Q8FYT2;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   GcpE protein.
GN   GCPE OR BR1778.
OS   Brucella suis.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC   Brucellaceae; Brucella.
OX   NCBI_TaxID=29461;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1330 / Biovar 1;
RX   MEDLINE=22247741; PubMed=12271122;
RA   Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,
RA   Read T.D., Dodson R.J., Umayam L., Brinkac L.M., Beanan M.J.,
RA   Daugherty S.C., Deboy R.T., Durkin A.S., Kolonay J.F., Madupu R.,
RA   Nelson W.C., Ayodeji B., Kraul M., Shetty J., Malek J., Van Aken S.E.,
RA   Riedmuller S., Tettelin H., Gill S.R., White O., Salzberg S.L.,
RA   Hoover D.L., Lindler L.E., Halling S.M., Boyle S.M., Fraser C.M.;
RT   "The Brucella suis genome reveals fundamental similarities between
RT   animal and plant pathogens and symbionts.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).
```

```
DR   EMBL; AE014468; AAN30676.1; -.
DR   TIGR; BR1778; -.
KW   Complete proteome.
SQ   SEQUENCE   420 AA;  45029 MW;  F1495DB2D73E164C CRC64;

  Query Match             27.2%;  Score 508;  DB 16;  Length 420;
  Best Local Similarity   33.2%;  Pred. No. 2.2e-29;
  Matches  128;  Conservative   79;  Mismatches  109;  Indels   70;  Gaps   13;

Qy          6 PIQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVRVSVPT 65
              |  ||:|  : || | :|  ||: |||||||  | ||::|| |: || | |::|||::|
Db         12 PFPRRQSVGVSVGGVIVGGSAPVVVQSMTNTDTADVDSTVAQVAALHRAGSEIVRITVDR 71

Qy         66 MDAAEAFKLIKQQV-----NVPLVADIHF-------DY-RIALKVAEYGVDCLRINPGNI 112
               ::| |   |::::     :|||| | |:       |:   |  :|:|     ||||||:
Db         72 DESAAAVPKIRERLERLGHDVPLVGDFHYIGHKLLADHPACAEALAKY-----RINPGNV 126

Qy        113 G----NEERIRMVVDCARDKNIPIRIGVNAGSLEKDL-------QEKYGEPTP------Q 155
              |     :::   :|: |   : |:||||| |||:::|       :  | |  :
Db        127 GFKDKKDKQFADIVEMAIRYDKPVRIGVNWGSLDQELLTTLMDRNQAEGAPLSAQDVMRE 186

Qy        156 ALLESAMRHVDHLDRLNF--DQFKVSVKASDVFLAVESYRLLAKQIDQPLHLGITEAGGA 213
              |:::||:   :  : :    |:  :| | | |   :  | :||:: :  ||||:||||
Db        187 AIVQSALISANLAEEIGLGRDKIILSAKVSQVQDLIAVYTMLAQRSNHALHLGLTEAGMG 246

Qy        214 RSGAVKSAIGLGLLLSEGIGDTLRVSLAADP----VEEIKVGFDILKSLRIRSRGINFIA 269
                | | |:  :|:|| :|||||:|:||  :|      |::|  ::|:::  |       |
Db        247 TKGIVASSAAMGILLQQGIGDTIRISLTPEPGGDRTREVQVAQELLQTMGFRQFVPIVAA 306

Qy        270 CPTCSRQEFDVIGTVNALEQRL-----EDI------------ITPMDVSIIGCVVNGPG 311
              || | |       | : : | |     |||            :  : |:::||:|||||
Db        307 CPGCGR-------TTSTVFQELAQTIQEDIRRNMPLWREKYPGVEALSVAVMGCIVNGPG 359

Qy        312 EALVSTLGV----TGGNKKSGLYEDG 333
              |:  : :|:    ||    : :: ||
Db        360 ESKHADIGISLPGTGETPSAPVFVDG 385


RESULT 10
Q8GZR6
ID   Q8GZR6      PRELIMINARY;      PRT;   740 AA.
AC   Q8GZR6;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   GcpE.
OS   Lycopersicon esculentum (Tomato).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; lamiids; Solanales; Solanaceae; Solanum.
OX   NCBI_TaxID=4081;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Querol J., Imperial S., Boronat A., Rodriguez-Concepcion M.;
RT   "Identification and functional analysis of plant orthologs of the
RT   Escherichia coli gcpE gene product with a role in plastid isoprenoid
RT   biosynthesis.";
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF435086; AAO15447.1; -.
SQ   SEQUENCE   740 AA;  82230 MW;  8668C9F08078FA05 CRC64;

  Query Match             25.2%;  Score 471;  DB 10;  Length 740;
  Best Local Similarity   40.1%;  Pred. No. 2.7e-26;
  Matches  109;  Conservative   45;  Mismatches   84;  Indels   34;  Gaps    5;

Qy          9 RRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVRVSVPTMDA 68
              ||::  : |||| :|   || :|:|| | | || ||| |:  :   |||||::|
Db         84 RRQTCTVMVGNVALGSEHPIRIQTMTTTDTKDVAATVEQVMKIADAGADIVRITVQGRKE 143

Qy         69 AEAF-----KLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGN---------.114
              |:|       |:::  |:|||||||||   :||:|||   | :|:|||| :
Db        144 ADACFEIKNTLVQKNYNIPLVADIHFAPSVALRVAEC-FDKIRVNPGNFADRRAQFEQLE 202

Qy        115 ------------EERIRMVVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLESA 161
                         ||    :|:  :     :||| | |||    :   ||: :|: ::|||
Db        203 YTEDDYQKELEHIEEVFTPLVEKCKKYGRAMRIGTNHGSLSDRIMSYYGD-SPRGMVESA 261

Qy        162 MRHVDHLDRLNFDQFKVSVKASDVFLAVESYRLLAKQI-----DQPLHLGITEAGGARSG 216
                      :|:|  |  |:|||:  : |::||||  ::      | |||||:||||    |
Db        262 FEFARICRKLDFHNFVFSMKASNPVVMVQAYRLLVAEMYVQGWDYPLHLGVTEAGEGEDG 321

Qy        217 AVKSAIGLGLLLSEGIGDTLRVSLAADPVEEI 248
               :||||||:| || :|:|||:||||    | |||
Db        322 RMKSAIGIGTLLQDGLGDTIRVSLTEAPEEEI 353


RESULT 11
Q8RXG8
ID   Q8RXG8      PRELIMINARY;      PRT;   741 AA.
AC   Q8RXG8;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   GcpE protein.
GN   AT5G60600.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,
RA   Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA   Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA   Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA   Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA   Ecker J., Theologis A., Davis R.W.;
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY081261; AAL91150.1; -.
DR   InterPro; IPR006705; GcpE.
DR   InterPro; IPR004588; IspG.
DR   Pfam; PF04551; GcpE; 1.
DR   TIGRFAMs; TIGR00612; ispG_gcpE; 1.
SQ   SEQUENCE   741 AA;  82157 MW;  957B11538F481585 CRC64;

  Query Match             24.8%;  Score 463;  DB 10;  Length 741;
  Best Local Similarity   38.6%;  Pred. No. 1e-25;
  Matches  105;  Conservative   47;  Mismatches   86;  Indels   34;  Gaps    5;

Qy          9 RRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVRVSVPTMDA 68
              |||:  : |||| :|   || :|:|| : | |:  ||::  :   ||||||::|
Db         84 RRKTRTVMVGNVALGSEHPIGIQTMTTSDTKDITGTVDEVMRIADKGADIVRITVQGKKE 143

Qy         69 AEAF-----KLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRMVVD 123
              |:|       ||::   |:|||||||||   :||:|||   | :|:|||| :       :|
Db        144 ADACFEIKDKLVQLNYNIPLVADIHFAPTVALRVAEC-FDKIRVNPGNFADRRAQFETID 202

Qy        124 CARDK---------------------NIPIRIGVNAGSLEKDLQEKYGEPTPQALLESA 161
                 |:                     :|||  | |||    :   ||: :|: ::|||
Db        203 YTEDEYQKELQHIEQVFTPLVEKCKKYGRAMRIGTNHGSLSDRIMSYYGD-SPRGMVESA 261

Qy        162 MRHVDHLDRLNFDQFKVSVKASDVFLAVESYRLLAKQI-----DQPLHLGITEAGGARSG 216
                      :|::  |  |:|||:  : |::||||  ::      |'|||||:||||    |
Db        262 FEFARICRKLDYHNFVFSMKASNPVIMVQAYRLLVAEMYVHGWDYPLHLGVTEAGEGEDG 321

Qy        217 AVKSAIGLGLLLSEGIGDTLRVSLAADPVEEI 248
               :|||||:| || :|:|||:||||    | |||
Db        322 RMKSAIGIGTLLQDGLGDTIRVSLTEPPEEEI 353


RESULT 12
Q9FF59
ID   Q9FF59      PRELIMINARY;      PRT;   716 AA.
AC   Q9FF59;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
```

```
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   GcpE protein.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Columbia;
RX   MEDLINE=97471969; PubMed=9330910;
RA   Sato S., Kotani H., Nakamura Y., Kaneko T., Asamizu E., Fukami M.,
RA   Miyajima N., Tabata S.;
RT   "Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence
RT   features of the 1.6 Mb regions covered by twenty physically assigned
RT   P1 clones.";
RL   DNA Res. 4:215-230(1997).
DR   EMBL; AB005246; BAB09833.1; -.
DR   InterPro; IPR004588; IspG.
DR   TIGRFAMs; TIGR00612; ispG_gcpE; 1.
SQ   SEQUENCE   716 AA;  79817 MW;  28D36FC644EB7CAA CRC64;

  Query Match           24.8%;  Score 462;  DB 10;  Length 716;
  Best Local Similarity   38.6%;  Pred. No. 1.2e-25;
  Matches  105;  Conservative   47;  Mismatches   86;  Indels   34;  Gaps    5;

Qy          9 RRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVRVSVPTMDA 68
              |||:  : |||| :|  || :|:|| : | |:  ||::  :   |||||::|
Db         59 RRKTRTVMVGNVALGSEHPIRIQTMTTSDTKDITGTVDEVMRIADKGADIVRITVQGKKE 118

Qy         69 AEAF-----KLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRMVVD 123
              |:|      ||::   |:|||||||   :||:|||   | :|:|||  :       :|
Db        119 ADACFEIKDKLVQLNYNIPLVADIHFAPTVALRVAEC-FDKIRVNPGNFADRRAQFETID 177

Qy        124 CARDK----------NIPIRIGVNAGSLEKDLQEKYGEPTPQALLESA 161
                |:                    :||| | |||   :   ||: :|: ::|||
Db        178 YTEDEYQKELQHIEQVFTPLVEKCKKYGRAMRIGTNHGSLSDRIMSYYGD-SPRGMVESA 236

Qy        162 MRHVDHLDRLNFDQFKVSVKASDVFLAVESYRLLAKQI-----DQPLHLGITEAGGARSG 216
                :|::  |  |:|||:  : |::|||||  ::     | |||||:||||    |
Db        237 FEFARICRKLDYHNFVFSMKASNPVIMVQAYRLLVAEMYVHGWDYPLHLGVTEAGEGEDG 296

Qy        217 AVKSAIGLGLLLSEGIGDTLRVSLAADPVEEI 248
               :|||||:| || :|:|||:||||   | |||
Db        297 RMKSAIGIGTLLQDGLGDTIRVSLTEPPEEEI 328


RESULT 13
Q8GZR7
ID   Q8GZR7      PRELIMINARY;      PRT;   740 AA.
AC   Q8GZR7;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   GcpE.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Querol J., Campos N., Imperial S., Boronat A.,
RA   Rodriguez-Concepcion M.;
RT   "Identification and functional analysis of plant orthologs of the
RT   Escherichia coli gcpE gene product with a role in plastid isoprenoid
RT   biosynthesis.";
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF434673; AAO15446.1; -.
SQ   SEQUENCE   740 AA;  82157 MW;  8EF625E9A9C88074 CRC64;

  Query Match           24.8%;  Score 462;  DB 10;  Length 740;
  Best Local Similarity   38.6%;  Pred. No. 1.2e-25;
  Matches  105;  Conservative   47;  Mismatches   86;  Indels   34;  Gaps    5;

Qy          9 RRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVRVSVPTMDA 68
              |||:  : |||| :|  || :|:|| : | |:  ||::  :   |||||::|
Db         84 RRKTRTVMVGNVALGSEHPIRIQTMTTSDTKDITGTVDEVMRIADKGADIVRITVQGKKE 143

Qy         69 AEAF-----KLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRMVVD 123
              |:|      ||::   |:|||||||   :||:|||   | :|:|||  :       :|
Db        144 ADACFEIKDKLVQLNYNIPLVADIHFAPTVALRVAEC-FDKIRVNPGNFADRRAQFETID 202

Qy        124 CARDK----------NIPIRIGVNAGSLEKDLQEKYGEPTPQALLESA 161
                |:                    :||| | |||   :   ||: :|: ::|||
Db        203 YTEDEYQKELQHIEQVFTPLVEKCKKYGRAMRIGTNHGSLSDRIMSYYGD-SPRGMVESA 261

Qy        162 MRHVDHLDRLNFDQFKVSVKASDVFLAVESYRLLAKQI-----DQPLHLGITEAGGARSG 216
                :|::  |  |:|||:  : |::|||||  ::     | |||||:||||    |
Db        262 FEFARICRKLDYHNFVFSMKASNPVIMVQAYRLLVAEMYVHGWDYPLHLGVTEAGEGEDG 321

Qy        217 AVKSAIGLGLLLSEGIGDTLRVSLAADPVEEI 248
               :|||||:| || :|:|||:||||   | |||
Db        322 RMKSAIGIGTLLQDGLGDTIRVSLTEPPEEEI 353


RESULT 14
Q8LPQ4
ID   Q8LPQ4      PRELIMINARY;      PRT;   741 AA.
AC   Q8LPQ4;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   AT5g60600/mup24_10.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Shinn P., Chen H., Cheuk R., Kim C.J., Meyers M.C., Banh J.,
RA   Bowser L., Carninci P., Chang E., Dale J.M., Goldsmith A.D.,
RA   Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
RA   Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M.,
RA   Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
RA   Seki M., Southwick A., Tang C.C., Toriumi M., Wu H.C., Yamada K.,
RA   Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA   Ecker J.R.;
RT   "Arabidopsis cDNA clones.";
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY094472; AAM19840.1; -.
DR   InterPro; IPR006705; GcpE.
DR   InterPro; IPR004588; IspG.
DR   Pfam; PF04551; GcpE; 2.
DR   TIGRFAMs; TIGR00612; ispG_gcpE; 1.
SQ   SEQUENCE   741 AA;  82257 MW;  BCA1D3147BD63ACB CRC64;

  Query Match           24.5%;  Score 457;  DB 10;  Length 741;
  Best Local Similarity   38.2%;  Pred. No. 2.9e-25;
  Matches  104;  Conservative   48;  Mismatches   86;  Indels   34;  Gaps    5;

Qy          9 RRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVRVSVPTMDA 68
              |||:  : ||:| :|  || :|:|| : | |:  ||::  :   |||||::|
Db         84 RRKTRTVMVGDVALGSEHPIRIQTMTTSDTKDITGTVDEVMRIADKGADIVRITVQGKKE 143

Qy         69 AEAF-----KLIKQQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRMVVD 123
              |:|      ||::   |:|||||||   :||:|||   | :|:|||  :       :|
Db        144 ADACFEIKDKLVQLNYNIPLVADIHFAPTVALRVAEC-FDKIRVNPGNFADRRAQFETID 202

Qy        124 CARDK----------NIPIRIGVNAGSLEKDLQEKYGEPTPQALLESA 161
                |:                    :||| | |||   :   ||: :|: ::|||
Db        203 YTEDEYQKELQHIEQVFTPLVEKCKKYGRAMRIGTNHGSLSDRIMSYYGD-SPRGMVESA 261

Qy        162 MRHVDHLDRLNFDQFKVSVKASDVFLAVESYRLLAKQI-----DQPLHLGITEAGGARSG 216
```

```
                    :|::  |  |:|||:  : |::||||  ::     | |||||:||||    |
Db        262 FEFARICRKLDYHNFVFSMKASNPVIMVQAYRLLVAEMYVHGWDYPLHLGVTEAGEGEDG 321

Qy        217 AVKSAIGLGLLLSEGIGDTLRVSLAADPVEEI 248
              :||||||:| ||  :|:|||:||||   | |||
Db        322 RMKSAIGIGTLLQDGLGDTIRVSLTEPPEEEI 353


RESULT 15
Q8KG23
ID   Q8KG23      PRELIMINARY;      PRT;   746 AA.
AC   Q8KG23;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   GcpE protein.
GN   GCPE OR CT0147.
OS   Chlorobium tepidum.
OC   Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae;
OC   Chlorobium.
OX   NCBI_TaxID=1097;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TLS / ATCC 49652 / DSM 12025;
RX   MEDLINE=22103685; PubMed=12093901;
RA   Eisen J.A., Nelson K.E., Paulsen I.T., Heidelberg J.F., Wu M.,
RA   Dodson R.J., Deboy R., Gwinn M.L., Nelson W.C., Haft D.H.,
RA   Hickey E.K., Peterson J.D., Durkin A.S., Kolonay J.L., Yang F.,
RA   Holt I., Umayam L.A., Mason T., Brenner M., Shea T.P., Parksey D.,
RA   Nierman W.C., Feldblyum T.V., Hansen C.L., Craven M.B., Radune D.,
RA   Vamathevan J., Khouri H., White O., Gruber T.M., Ketchum K.A.,
RA   Venter J.C., Tettelin H., Bryant D.A., Fraser C.M.;
RT   "The complete genome sequence of Chlorobium tepidum TLS, a
RT   photosynthetic, anaerobic, green-sulfur bacterium.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:9509-9514(2002).
DR   EMBL; AE012794; AAM71395.1; -.
DR   TIGR; CT0147; -.
DR   InterPro; IPR006705; GcpE.
DR   InterPro; IPR004588; IspG.
DR   Pfam; PF04551; GcpE; 2.
DR   TIGRFAMs; TIGR00612; ispG_gcpE; 1.
KW   Complete proteome.
SQ   SEQUENCE   746 AA;  82087 MW;  B40ED95B6010C5F9 CRC64;

  Query Match           24.4%;  Score 455.5;  DB 16;  Length 746;
  Best Local Similarity   37.5%;  Pred. No. 3.8e-25;
  Matches  108;  Conservative   51;  Mismatches   90;  Indels   39;  Gaps    8;

Qy          5 API---QRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVRV 61
              ||:   :|| :  :  | : :|   || |:||    | |  |:| | : |   | :|:|:
Db         68 APVYSYRRRVTREVPFGTIFLGGYLPIRVESMITAHTMDTAASVEQCRRLYEAGCEIIRL 127

Qy         62 SVPTMDAAEAFKLIKQQ-----VNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGN-- 114
              :|||   ||  | |:|     :: |||||||  : |:|  |: |: :||||||
Db        128 TVPTEKDAENLKNIREQLRRDGIDTPLVADIHFSAKAAMKAVEF-VENIRINPGNYATGA 186

Qy        115 ------------------EERIRMVVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQ 155
                               |    :|  ||   : :||| | |||   :  :||  :|:
Db.       187 KFSSKDYTDDEYRAELDKVREEFTPLVRKARSLGVSMRIGTNHGSLSDRIVSRYGN-SPE 245

Qy        156 ALLESAMRHVDHL-DRLNFDQFKVSVKASDVFLAVESYRLLAKQID------QPLHLGIT 208
              ::|:|:       |   :||   |:|:|:| : :::|||| | :|       ||||:|
Db        246 GMVEAALEFSRICEDEGYYDQL-FSMKSSNVRVMIQAYRLLVARADAELRYAYPLHLGVT 304

Qy        209 EAGGARSGAVKSAIGLGLLLSEGIGDTLRVSLAADPVEEIKVGFDILK 256
              |||    | :|||:|:| || :|:|||:|||| ||| |: ||| |:|
Db        305 EAGDGDEGRIKSAMGIGALLEDGLGDTIRVSLTEDPVNEVPVGFAIVK 352


Search completed: August 12, 2003, 09:58:37
Job time : 102 secs
```